②

815/6.870

OF EDNL. & PSYL. LIBRARY
2

Bureau Edmk Psy Research
[illegible] COLLEGE
Dated ....
No ....

73

BIOMETRIKA

BIOMETRIKA

FOUNDED BY
W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

MANAGING EDITOR
E. S. PEARSON

ASSOCIATE EDITORS
M. G. KENDALL JOHN WISHART

in consultation with
HARALD CRAMÉR R. C. GEARY
J. B. S. HALDANE

VOLUME 40

1953

BUREAU OF EDNL. & PSYL. RESEARCH
LIBRARY

ISSUED BY
THE BIOMETRIKA OFFICE, UNIVERSITY COLLEGE, LONDON
PRINTED AT THE UNIVERSITY PRESS, CAMBRIDGE

[illegible] Psy. Research
[illegible] TRAINING COLLEGE
6. 8. 70
J. 15.

PRINTED IN GREAT BRITAIN

CONTENTS OF VOLUME *40*

Memoirs

PAGES

I. The superposition of several strictly periodic sequences of events. By D. R. Cox and W. L. Smith. With four Figures in the Text 1—11

II. Approximate confidence intervals. By M. S. Bartlett . . 12—19

III. Incomplete and absolute moments of the multivariate normal distribution with some applications. By A. R. Kamat . . 20—34

IV. On the range of partial sums of a finite number of independent normal variates. By A. A. Anis and E. H. Lloyd . . . 35—42

V. Note on 'the Jacobians of certain matrix transformations useful in multivariate analysis'. By Ingram Olkin . . . . 43—46

VI. Estimation of a functional relationship. By D. V. Lindley . 47—49

VII. Estimating parameters in truncated Pearson frequency distributions without resort to higher moments. By A. C. Cohen, Jr. 50—57

VIII. A problem of interference between two queues. By J. C. Tanner 58—69

IX. Tables of the angular transformation. By W. L. Stevens . . 70—73

X. Tests of significance in a 2 × 2 contingency table: extension of Finney's table. Computed by R. Latscha . . . . . 74—86

XI. A method for judging all contrasts in the analysis of variance. By Henry Scheffe . . . . 87—104

XII. The estimation and comparison of strengths of association in contingency tables. By A. Stuart . . . . 105—110

XIII. A sequential test for randomness. By P. G. Moore. With one Figure in the Text . . . . 111—115

XIV. On the mean successive difference and its ratio to the root mean square. By A. R. Kamat . . . . 116—127

XV. The effect of unequal group variances on the F-test for the homogeneity of group means. By G. Horsnell. With two Figures in the Text . . . . 128—136

XVI. The estimation of population parameters from data obtained by means of the capture-recapture methods. III. An example of the practical applications of the method. By P. H. Leslie, Dennis Chitty and Helen Chitty. With one Figure in the Text . . . . 137—169

XVII. On the utilization of marked specimens in estimating populations of flying insects. By C. C. Craig . . . . . . 170—176

PAGES

XVIII. The total size of a general stochastic epidemic. By NORMAN T. J. BAILEY. With three Figures in the Text . . . . . . 177—185

XIX. Experimental evidence concerning contagious distributions in ecology. By D. A. EVANS. With four Figures in the Text . 186—211

XX. The population frequencies of species and the estimation of population parameters. By I. J. GOOD . . . . . . 237—264

XXI. Capture-recapture analysis. By J. M. HAMMERSLEY . . . 265—278

XXII. The use of chain-binomials with a variable chance of infection for the analysis of intra-household epidemics. By NORMAN T. J. BAILEY . . . . 279—286

XXIII. Spread of diseases in a rectangular plantation with vacancies. By G. H. FREEMAN. With two Figures in the Text . . . 287—296

XXIV. Tests of significance for concurrent regression lines. By E. J. WILLIAMS . . . . 297—305

XXV. Approximate confidence intervals. II. More than one unknown parameter. By M. S. BARTLETT . . . . . . 306—317

XXVI. Non-normality and tests on variances. By G. E. P. BOX. With two Figures in the Text . . . . 318—335

XXVII. Approximating to the distributions of measures of dispersion by a power of χ^2. By J. H. CADWELL . . . . . . 336—346

XXVIII. The power function of some tests based on range. By H. A. DAVID 347—353

XXIX. Some simple approximate tests for Poisson variates. By D. R. COX. . . . . 354—360

XXX. Orthogonal polynomial fitting. By JOHN WISHART and THEOCHARIS METAKIDES . . . . 361—369

XXXI. Population differences between species growing according to simple birth and death processes. By J. H. DARWIN . . 370—382

XXXII. Modifications to the variate-difference method. By M. H. QUENOUILLE . . . . 383—408

XXXIII. Moments of the rank correlation coefficient τ in the general case. By R. M. SUNDRUM . . . . 409—420

XXXIV. 99·9 % and 0·1 % points of the χ^2-distribution. By T. LEWIS . 421—426

XXXV. Tables of symmetric functions. Part IV. By F. N. DAVID and M. G. KENDALL . . . . 427—446

Miscellanea

PAGES

(i) Time intervals between accidents—a note on Maguire, Pearson and Wynn's paper. By G. A. BARNARD . . . . 212—213

(ii) Further notes on the analysis of accident data. By B. A. MAGUIRE, E. S. PEARSON and A. H. A. WYNN. With one Figure in the Text 214—216

(iii) On a method of estimating biological populations in the field. By C. C. CRAIG. With one Figure in the Text . . . . . . 216—218

(iv) A rapid method for estimating the correlation coefficient from the range of the deviations about the reduced major axis. By C. H. LEIGH-DUGMORE . . . . 218—219

(v) The effect of overlapping in bacterial counts of incubated colonies. By C. MACK . . . . 220—222

(vi) Non-normality in two-sample *t*-tests. By D. G. C. GRONOW . . 223—225

(vii) Note on the Poisson Index of Dispersion. By N. KATHIRGAMATAMBY 225—228

(viii) On an extension of Geary's theorem. By R. G. LAHA . . . 228—229

(ix) The Doolittle method and the fitting of polynomials to weighted data. By P. G. GUEST . . . . 229—231

(x) A simple method of deriving best critical regions similar to the sample space in tests of an important class of composite hypotheses. By K. S. RAO . . . . 231—233

(xi) Review of Gerhard Tintner's *Econometrics*. By J. DURBIN . . 234

(xii) Review of Acheson J. Duncan's *Quality Control and Industrial Statistics*. By D. R. COX . . . . 234—235

(xiii) Some procedures for comparing Poisson processes or populations. By ALLAN BIRNBAUM . . . . 447—449

(xiv) Scale factors and degrees of freedom for small sample sizes for χ-approximation to the range. By GEORGE WM. THOMSON . . 449—450

(xv) The third moment of Gini's mean difference. By A. R. KAMAT . 451—452

(xvi) A method of systematic sampling based on order properties. By R. M. SUNDRUM . . . . 452—456

(xvii) A note on ordered least-squares estimation. By F. DOWNTON . 457—458

(xviii) A note on the evaluation of the multivariate normal integral. By F. N. DAVID . . . . 458—459

(xix) A graphical method for the analysis of statistical distributions into two normal components. By ERIC J. PRESTON. With one Figure in the Text . . . . 460—464

(xx) A note on regions for tests of kurtosis. By G. E. P. BOX. With two Figures in the Text . . . . 465—468

PAGES

(xxi) The frequency justification of sequential tests—addendum. By G. A. BARNARD . . . . 468—469

(xxii) Review of O. Kempthorne's *The Design and Analysis of Experiments.* By C. A. MOSER . . . . 470—471

(xxiii) Review of M. H. Quenouille's *Associated Measurements* and *The Design and Analysis of Experiment.* By D. R. COX . . . . 471—472

(xxiv) Review of C. I. Bliss's *The Statistics of Bio-assay* and D. J. Finney's *Statistical Method in Biological Assay* and *Probit Analysis.* By M. J. R. HEALY . . . . , . 473

(xxv) Review of A. Hald's *Statistical Theory with Engineering Applications* and *Statistical Tables and Formulas.* By P. G. MOORE . . . 473—474

(xxvi) Review of A. Wald's *Statistical Decision Functions.* By G. A. BARNARD . . . . 475—477

(xxvii) Review of *Applied Mathematics Series*, Nos. 11, 14, 21, 25 and 26, of the National Bureau of Standards, U.S. Department of Commerce. By H. O. HARTLEY . . . . 477—478

VOLUME 40, PARTS 1 AND 2 JUNE 1953

THE SUPERPOSITION OF SEVERAL STRICTLY PERIODIC SEQUENCES OF EVENTS

BY D. R. COX AND W. L. SMITH

Statistical Laboratory, University of Cambridge

1. INTRODUCTION

Suppose that there are a number of *sources* at each of which *events* occur from time to time. Suppose further that the *outputs* of the various sources are combined into one *pooled output*. We shall consider the problem of deducing the statistical properties of the pooled output from the statistical properties of the separate outputs. By a statistical property we mean, for example, the frequency distribution of the time interval between successive events. The problem is illustrated for three sources in Fig. 1. The events are marked by short vertical lines and the outputs of the sources S_1, S_2 and S_3 are combined into a single pooled output P.

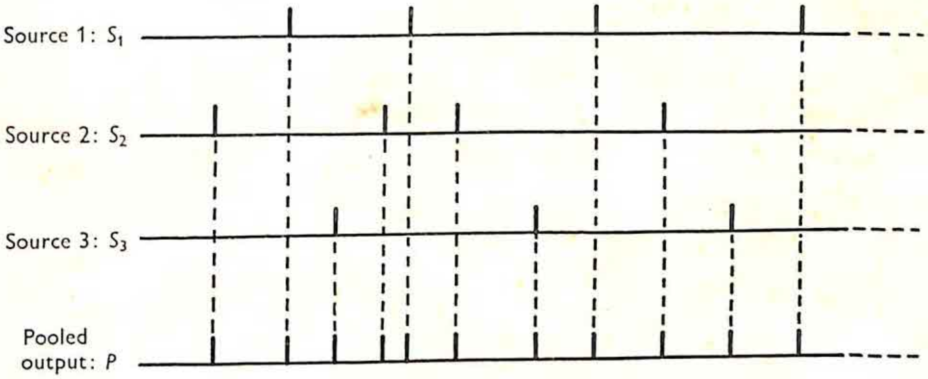


Fig. 1. The pooling of outputs.

If we denote by $t_j^{(i)}$ the time at which the jth event occurs on the ith source, we can state the problem formally as follows. Given N increasing sequences $S_i \equiv [t_1^{(i)}, t_2^{(i)}, \ldots]$ $(i = 1, \ldots, N)$, let the $t_j^{(i)}$ be rearranged into a single increasing sequence P. The statistical properties of the S_i are assumed known and we wish to determine various properties of the pooled output P.

There are two cases of special interest. First, the intervals between successive events on any one source may be independent identically distributed random variables, the random variables associated with different sources also being independent. This situation is closely connected with renewal theory and will be discussed in a separate paper (Cox & Smith, 1953). The second case, to which the present paper is devoted, arises when the events on any one source occur at exactly regular intervals so that the sequence S_i is $[\theta_i, 2\theta_i, \ldots]$, where θ_i is the *period* of the ith source.

At first sight this is not a statistical problem at all, because the times at which events occur are completely determined and no probability or chance mechanism enters the system. That statistical concepts are in fact relevant follows from a well-known theorem in the theory of numbers due to H. Weyl. This states, in effect, that if the θ_i are mutually irrational, the sources are random with respect to one another if observed after a sufficiently long time. We make this remark precise in § 2.

Two applications, one in experimental psychology and the other in neuro-physiology, will now be described.

Conrad (1951) has investigated behaviour in situations calling for constant vigilance and for adaptive reactions to a changing pattern of events. He used a number of dials, on each of which a pointer revolved at constant speed, different for different dials. Every time a pointer passed a mark on its dial the operator was required to turn a knob. Conrad found how the number of responses that the operator omitted to make and the error of timing depended on the number and speed of the dials. Here the sequence of events on any one dial is periodic, and the pooled output is the sequence of instants at which response is called for from the operator. The behaviour of the operator is likely to depend both on the mean number of events per minute and on the number of times events occur very close together. It is therefore important to know the frequency distribution of the interval between successive events in the pooled output.*

One neurological application arose in the recent work by Fatt & Katz (1952) on spontaneous subthreshold activity at motor-nerve endings. They found that in certain many-branching nerve endings there are a large number of active spots each passing electrical disturbances into muscle fibre. Fatt & Katz showed that the frequency distribution of the time interval between successive disturbances in the muscle fibre agreed with the theoretical exponential distribution characteristic of a random series. However, they explicitly pointed out that the succession of pulses from a single spot might nevertheless be periodic. The statistical problem here is to infer as much as possible about the individual sources (i.e. the single active spots) from observations on the pooled output. Another application is obtained when the sources are neurons and the events are discrete nerve pulses sent to some central nerve cell. This possibility is relevant to the theory of neural nets developed by Householder, Landahl and others (Householder & Landahl, 1945).

2. FREQUENCY DISTRIBUTION OF INTERVALS IN THE POOLED OUTPUT

In this section we show how Weyl's theorem (Weyl, 1916)† can be used to derive the frequency distribution of the interval between successive events in the pooled output. It will be assumed that the periods θ_i are positive numbers, and are *mutually irrational* in the sense that there exists no set of positive or negative integers n_i, not all zero, such that

$$\sum_{i=1}^{N} n_i \theta_i = 0. \tag{1}$$

In particular, the θ_i are all different and θ_i/θ_j is irrational for all $i \neq j$. We suppose that the sources are numbered so that θ_N is the smallest of the θ_i.

The simple form of Weyl's theorem states that if α is irrational, and if $\{x\}$ denotes the fractional part of x, then the sequence $[\{n\theta\}]$ $(n = 1, 2, \ldots)$ is uniformly distributed over $(0, 1)$. More precisely, if I_l is any interval of length l in $(0, 1)$ and if $p_m(I_l)$ is the proportion of $\{\theta\}, \{2\theta\}, \ldots, \{m\theta\}$ falling in I_l, then

$$\lim_{m \to \infty} p_m(I_l) = l. \tag{2}$$

---

* We are grateful to Mr Conrad for suggesting the investigation of this distribution.

† This, though not the earliest or most accessible reference, contains a particularly elegant proof of the theorem both in its simple and in its generalized form.

In fact, all the frequency distributions considered in this paper are defined in an analogous way by limiting frequencies in a precisely defined infinite sequence. The successive members of the sequence are in no sense independent or random, and hence we have refrained from using the term probability.

To reformulate (2) in terms of sources, consider any two sources, say the first two. Associate with the rth event on the first source a quantity x_r equal to the time between that event and the immediately preceding event on the second source (see Fig. 2). We have

$$x_r = \theta_2\{r\theta_1/\theta_2\}. \tag{3}$$

It follows from Weyl's theorem that, since θ_1/θ_2 is irrational, the sequence $[x_r]$ is uniformly distributed over $(0, \theta_2)$.

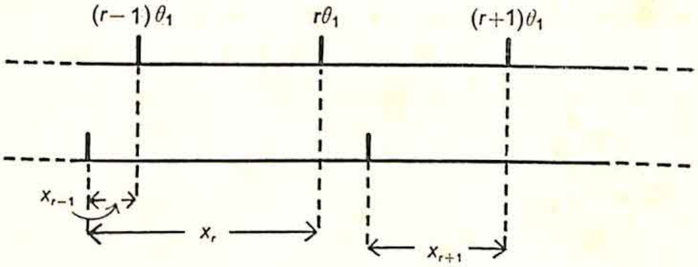


Fig. 2. Definition of x_r.

The generalized form of Weyl's theorem states that if $\alpha_1, \ldots, \alpha_k$ are irrational numbers, themselves mutually irrational in the sense of (1), then the sequences $[\{n\alpha_1\}], \ldots, [\{n\alpha_k\}]$ are *independently* uniformly distributed over $(0, 1)$. More precisely, if I_v is any portion of volume v of the k-dimensional unit cube, and if $p_m(I_v)$ is the proportion of the m k-plets

$$\{\alpha_1\}, \ldots, \{\alpha_k\}; \;\ldots;\; \{m\alpha_1\}, \ldots, \{m\alpha_k\}$$

in I_v, then

$$\lim_{m\to\infty} p_m(I_v) = v. \tag{4}$$

To apply this to our problem we define for the rth event on, say, the ith source a set of $(N-1)$ quantities $x_r^{(1)}, \ldots, x_r^{(i-1)}, x_r^{(i+1)}, \ldots, x_r^{(N)}$ analogous to x_r in (3). For example, $x_r^{(l)}$ is the interval between the rth event on the ith source and the immediately preceding event on the lth source. Then (4) shows that the sequences $[x_r^{(1)}], \ldots, [x_r^{(i-1)}], [x_r^{(i+1)}], \ldots, [x_r^{(N)}]$ are independently uniformly distributed over $(0, \theta_1), \ldots, (0, \theta_N)$.

We can now calculate the frequency distribution of the interval between successive events in the pooled output. For let $q_i(y)$ be the frequency function of the length of intervals ending with an event from source i. The interval is between $(y, y+\delta y)$ if one source, say the jth, has its x-value between $(y, y+\delta y)$ and if all the other sources have their x-values greater than y. The frequency of the first of these events is $\delta y/\theta_j$, and of the second $\prod_{k\neq i,j} (\theta_k - y)/\theta_k$. Thus, by the independence of the x-values,

$$q_i(y) = \sum_{j\neq i} \frac{1}{\theta_j} \prod_{k\neq i,j} \frac{(\theta_k - y)}{\theta_k} \quad (0 \leqslant y \leqslant \theta_N). \tag{5}$$

Also when $i = N$ there is a point concentration Q_N of frequency at $y = \theta_N$ given by

$$Q_N = \prod_{k=1}^{N-1} \frac{(\theta_k - \theta_N)}{\theta_k}. \tag{6}$$

If $i \neq N$ there is no point frequency.

The overall frequency distribution is defined by a frequency function $q(y)$, and by a point frequency Q, and is obtained by taking a weighted average of the $q_i(y)$. In a long time the number of events from the ith source is proportional to θ_i^{-1}, so that

$$q(y) = \sum_{i=1}^{N} \theta_i^{-1} q_i(y) \Big/ \sum_{i=1}^{N} \theta_i^{-1}$$

$$= \frac{\displaystyle\sum_{i=1}^{N} \frac{1}{\theta_i} \sum_{j \neq i} \frac{1}{\theta_j} \prod_{k \neq i,j} \frac{(\theta_k - y)}{\theta_k}}{\displaystyle\sum_{i=1}^{N} \frac{1}{\theta_i}} \quad (0 \leqslant y \leqslant \theta_N), \tag{7}$$

and

$$Q = \frac{\displaystyle\frac{1}{\theta_N} \prod_{k=1}^{N-1} \frac{(\theta_k - \theta_N)}{\theta_k}}{\displaystyle\sum_{i=1}^{N} \frac{1}{\theta_i}}. \tag{8}$$

We shall write (7) and (8) in two forms, one suitable for exact computation when N is small and the other suitable for deriving asymptotic forms when N is large.

To obtain an expression for computation, we introduce the symmetric functions

$$\left.\begin{aligned} \sigma_1^{(N)} &= \theta_1 + \ldots + \theta_N, \\ \sigma_2^{(N)} &= \sum_{i_1 > i_2} \theta_{i_1} \theta_{i_2}, \\ \sigma_s^{(N)} &= \sum_{i_1 > \ldots > i_s} \theta_{i_1} \ldots \theta_{i_s}, \\ \sigma_N^{(N)} &= \theta_1 \ldots \theta_N. \end{aligned}\right\} \tag{9}$$

Then

$$q(y) = \frac{1}{\sigma_{N-1}^{(N)}} \sum_{i=1}^{N} \sum_{j \neq 1} \prod_{k \neq i,j} (\theta_k - y)$$

$$= \frac{1}{\sigma_{N-1}^{(N)}} \sum_{r=0}^{N-2} (-1)^r (r+1)(r+2)\, y^r \sigma_{N-2-r}^{(N)} \quad (0 \leqslant y \leqslant \theta_N), \tag{10}$$

and

$$Q = \frac{\displaystyle\prod_{k=1}^{N-1} (\theta_k - \theta_N)}{\sigma_{N-1}^{(N)}}. \tag{11}$$

The coefficient of y^r in (10) is a sum of products of different θ's with $(r+2)$ missing, and hence by symmetry must be a multiple of $\sigma_{N-2-r}^{(N)}$. The numerical coefficient is obtained by considering the number of ways of permuting the suffices i, j, k.

Equations (10) and (11) define the frequency distribution of intervals on the pooled output. The distribution consists of a discrete frequency, Q, at the smallest period θ_N, and a continuous frequency curve between 0 and θ_N defined by a polynomial of degree $(N-2)$. When N is small (10) and (11) are easily computed numerically; some examples are given in § 5. The symmetric functions, $\sigma_r^{(N)}$, can be found systematically by a method of undetermined multipliers.

The mean of the distribution (10), (11) can be found from first principles. In a very long time T there are asymptotically $T\theta_i^{-1}$ events on the ith source, and hence asymptotically $T\Sigma\theta_i^{-1}$ events altogether. Therefore

$$E(y) = \left[\sum_{i=1}^{N} \theta_i^{-1}\right]^{-1}. \tag{12}$$

The method used to obtain the frequency distribution of the interval between events in the pooled output can be extended to give the joint frequency distribution of successive intervals.

3. ASYMPTOTIC PROPERTIES

In this section we first show that if the θ_i satisfy some weak conditions, the distribution (7) tends to the exponential distribution as N tends to infinity.

Let us set $\chi_i = \theta_i^{-1}$ and express y as a multiple of its mean by writing

$$z = y \sum_{i=1}^{N} \chi_i. \tag{13}$$

The frequency function, $r(z)$, of z is then from (7)

$$r(z) = \frac{\sum\limits_{i \neq j} \prod\limits_{k \neq i,j} \left(1 - \frac{z\chi_k}{\Sigma\chi_l}\right)\chi_i\chi_j}{(\Sigma\chi_l)^2} = \frac{d^2\rho(z)}{dz^2}, \tag{14}$$

where

$$\rho(z) = \prod_{i=1}^{N}\left(1 - \frac{z\chi_i}{\Sigma\chi_l}\right). \tag{15}$$

To discuss the limiting form of (15) we introduce the moments of the finite population $\chi_1, \ldots, \chi_N$. Put

$$\left.\begin{aligned} \mu &= (\Sigma\chi_i)/N, \\ \mu_s' &= (\Sigma\chi_i^s)/N \quad (s = 2, 3, \ldots). \end{aligned}\right\} \tag{16}$$

Then

$$\log \rho(z) = \sum_{i=1}^{N} \log\left(1 - \frac{z\chi_i}{\Sigma\chi_l}\right) = -z - \frac{z^2}{2}\frac{\mu_2'}{N\mu^2} - \frac{z^3}{3}\frac{\mu_3'}{N^2\mu^3} - \ldots. \tag{17}$$

To justify a limiting operation on (17) various assumptions about the χ_i are possible. They all have the effect of ensuring that no small group of periods is comparable with the mean interval (12) as N tends to infinity. The strongest assumption is:

(i*a*) As N tends to infinity the χ_i have both an upper bound and a non-zero lower bound. This implies:

(i*b*) As N tends to infinity the dimensionless moments μ_s'/μ^s are bounded.

A weaker assumption still is:

(i*c*) As N tends to infinity $\mu_s'/\mu^s \leqslant A^s N^{s-2}$ for some A, $s = 2, 3, \ldots$.

Under any of (i*a*), (i*b*) or (i*c*) we can let N tend to infinity in (17) and obtain

$$\rho(z) \sim e^{-z}, \tag{18}$$

and

$$r(z) = \frac{d^2\rho(z)}{dz^2} \sim e^{-z}. \tag{19}$$

The differentiation of (18) is justified because $\rho(z)$ for all finite N and the limit function e^{-z} are integral functions. If we retain the term in N^{-1}, we get under assumption (i*a*) or (i*b*) the second approximation

$$r_2(z) = e^{-z}\left[1 - \frac{(1+C^2)(z^2 - 4z + 2)}{2N}\right], \tag{20}$$

where $C^2 = \mu_2'/\mu^2 - 1$ is the squared coefficient of variation of the χ_i.

The continuous frequency curve, $r(z)$, has range $0 \leqslant z \leqslant \chi_N^{-1}\Sigma\chi_i$. We shall assume that the smallest period θ_N is large compared with the mean interval, i.e. that

$$\text{(ii)} \qquad \lim_{N\to\infty} \chi_N^{-1} \sum_{i=1}^{N} \chi_i = \infty, \tag{21}$$

and it can then be shown that the point frequency (8) tends to zero. Thus the continuous frequency curve accounts asymptotically for the whole of the frequency.

We have shown that under assumptions (i) and (ii) the frequency distribution of the interval between successive events on the pooled output tends to the exponential curve. It can be shown similarly that successive intervals tend to be distributed in independent exponential distributions. These results are but one aspect of what may be called the *local randomness* of the pooled output. As another illustration of local randomness we now prove that, if t is small compared with all the periods θ_i, the number of events observed in time t follows a Poisson distribution of mean $t(\Sigma\chi_i)$. The frequency distribution of the number of events in time t is defined by taking a large number of intervals of length t spread uniformly over a very long time, and finding the proportion of intervals containing $0, 1, 2, \ldots$ events. More precisely, the intervals should be of the form $(n\alpha, n\alpha+t)$ $(n = 1, 2, \ldots)$, where α is mutually irrational with $\theta_1, \ldots, \theta_N$.

Consider first a single source. A time interval $t<\theta_i$ will include either no event or one event, and the proportion of intervals including one event is t/θ_i. Thus the generating function of the frequencies is

$$(1-t/\theta_i)+\zeta t/\theta_i = 1-(1-\zeta)\,t\chi_i.$$

It follows from the independence of the sources that the generating function for the pooled output is

$$G(\zeta) = \prod_{i=1}^{N} [1-(1-\zeta)\,t\chi_i],$$

so that

$$\log G(\zeta) = \sum_{i=1}^{N} \log[1-(1-\zeta)\,t\chi_i].$$

If we assume that $t\chi_i$ is small for all i,

$$\log G(\zeta) \sim \sum_{i=1}^{N} -t\chi_i(1-\zeta),$$

i.e.

$$G(\zeta) \sim \exp[-t(1-\zeta)\,\Sigma\chi_i], \tag{22}$$

the generating function of the Poisson distribution.

The asymptotic equation (22) holds for times t small compared with the individual periods. If N is large, t may nevertheless contain an appreciable number of events on the pooled output. If N is small t must be smaller than, or of the order of, the mean interval between events and in this case the result is, in a sense, vacuous.

The results of this section are special cases of the following intuitively obvious principle. Suppose that there are a number of sources and that we observe some property of the pooled output depending only on its behaviour over times small compared with the individual recurrence times. Then, except in degenerate cases, the result is indistinguishable from that for a random series, no matter what the form of the individual outputs. In particular if the number of sources and the individual recurrence times are large the output will be random over intervals containing many events.

4. THE VARIANCE-TIME CURVE

The pooled output is distinguished from a random series by its behaviour over lengths of time comparable with the individual periods θ_i. The most convenient way of expressing this behaviour is by the *variance-time curve*, $V(t)$. This is defined as the variance of the number of events occurring in a time t, considered as a function of t. For a random series

$$V(t) = \lambda t, \tag{23}$$

where λ^{-1} is the mean interval between successive events. The analogue of $V(t)$ for continuous stochastic processes has been extensively used, for example in studies of sampling and in work on irregularity in textile processing.

To find $V(t)$ for the pooled output of periodic sources, consider first a single source. Let $\chi_i t = n_i + \beta_i$, where n_i is an integer and $0 \leqslant \beta_i < 1$, i.e.

$$\beta_i = \{\chi_i t\}.$$

Then an interval of length t contains either n_i or n_i+1 events from this source and the limiting frequency of intervals containing n_i+1 events is β_i. The variance of the two-point distribution is $\beta_i(1-\beta_i)$. Since the different sources are, by Weyl's theorem, independent, we have for N sources

$$V(t) = \sum_{i=1}^{N} \beta_i(1-\beta_i). \tag{24}$$

More generally the generating function of the frequency distribution of the number of events in an interval t is

$$G(\zeta) = \prod_{i=1}^{N} \zeta^{n_i}[1-\beta_i+\beta_i\zeta]. \tag{25}$$

The general behaviour of (24) can be seen as follows. If t is very small compared with the θ_i,

$$\beta_i(1-\beta_i) \sim \beta_i = t\chi_i$$

so that

$$V(t) \sim t\Sigma\chi_i = t\lambda, \tag{26}$$

where λ^{-1} is the mean interval between successive events. Further

$$\beta_i(1-\beta_i) < t\chi_i,$$

so that

$$V(t) < t\lambda. \tag{27}$$

For t large compared with θ_i,

$$\beta_i(1-\beta_i) \ll t\chi_i,$$

so that

$$V(t) \ll t\lambda.$$

Finally, as t increases, β_i takes each value between 0 and 1 equally often giving $\beta_i(1-\beta_i)$ an average value of $\frac{1}{6}$. Thus for large t, $V(t)$ oscillates about an average of $\frac{1}{6}N$.

To sum up, $V(t)$ is tangential at $t = 0$ to the straight line (23) representing a random series with the same mean number of events per unit time. $V(t)$ falls below (23) as soon as t is comparable with an appreciable number of the periods θ_i, and finally as t tends to infinity $V(t)$ oscillates about $\frac{1}{6}N$. Some numerical examples are given in § 5.

We have assumed above that N is finite and this is certainly true in the applications we have in mind. We now give a very brief discussion of the behaviour when N is infinite. If $\Sigma\chi_i$ is divergent the mean interval between successive events is zero so that time intervals can be found containing infinitely many events. If we exclude this possibility $\Sigma\chi_i$ must converge. By the argument for the finite case, $V(t)$ oscillates for large t about the curve $\frac{1}{6}N(t)$, where $N(t)$ is the number of periods less than t. As t tends to infinity, $N(t)$ tends to infinity but the rate of increase of $N(t)$ tends to zero by the convergence of $\Sigma\chi_i$.

5. NUMERICAL EXAMPLES

Three special cases with N small have been investigated numerically. They are

Example A: $N = 4$, Periods 1, 1·1317, 1·2781, 1·3486.

Example B: $N = 4$, Periods 1, 1·1317, 3·2781, 6·8914.

Example C: $N = 10$, Periods 1, 1·1327, 1·3819, 1·4662, 1·6173, 1·6593, 1·7814, 1·8181, 2·2203, 2·6421.

Fig. 3 shows the frequency distribution of the interval between successive events giving

(*a*) the exact polynomial (10);

(*b*) the exponential curve of the same mean;

(*c*) the second approximation (20).

Fig. 4 shows the variance-time curves computed from equation (25). The curve for Example C has approximately thirty cusps in the range shown; no attempt has been made to depict all these.

The following conclusions may be drawn:

(i) The second approximation (20) works well even when N is quite small.

(ii) For a given number of periods, the distribution of intervals is closest to an exponential distribution when the periods θ_i are nearly equal.

(iii) The distribution of intervals tends rapidly to the exponential form as N increases, if the θ_i are close together.

(iv) The variance-time curves have a characteristic form entirely different from that for a random series.

For computational convenience the θ_i could not be mutually irrational. We may either regard them as close approximations to irrationals or argue that since the lowest common multiple of the θ_i is extremely large, the results derived on the assumption (1) are likely to be extremely good approximations to the behaviour for rational θ_i.

6. ANALYSIS OF DATA

Suppose we are given a series of events suspected to be the pooled output of a number of periodic sources. What methods of analysis can be suggested?

If the number of sources is small and if the series available for analysis is long, it is possible in principle both to determine the θ_i exactly and to assign each event to its appropriate source. To do this, first form the frequency distribution of the intervals between successive events. This will be bounded by a point concentration of frequency whose position will determine the smallest period θ_N. Next find an interval of length θ_N and from it build up the output of the Nth source by repeated addition and subtraction of θ_N. Delete this set of events and then analyse the remaining events to find the next smallest period, and so on. The method ceases to be a practical one as soon as the point frequency Q becomes very small, and therefore it is most unlikely to be useful in the neurological applications where the number of sources may be one hundred or more. For this we need less direct methods which we now consider very briefly.

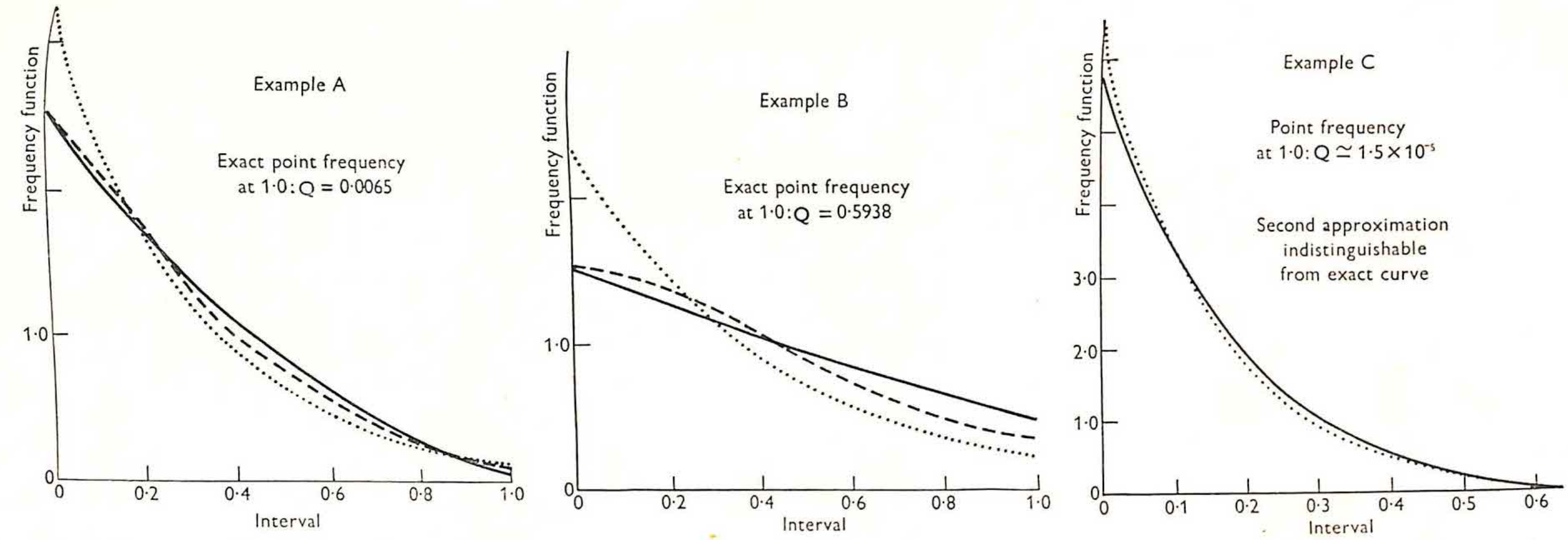


Fig. 3. Frequency distribution of interval between successive events. —— exact expression; ..... exponential curve; ----- second approximation.

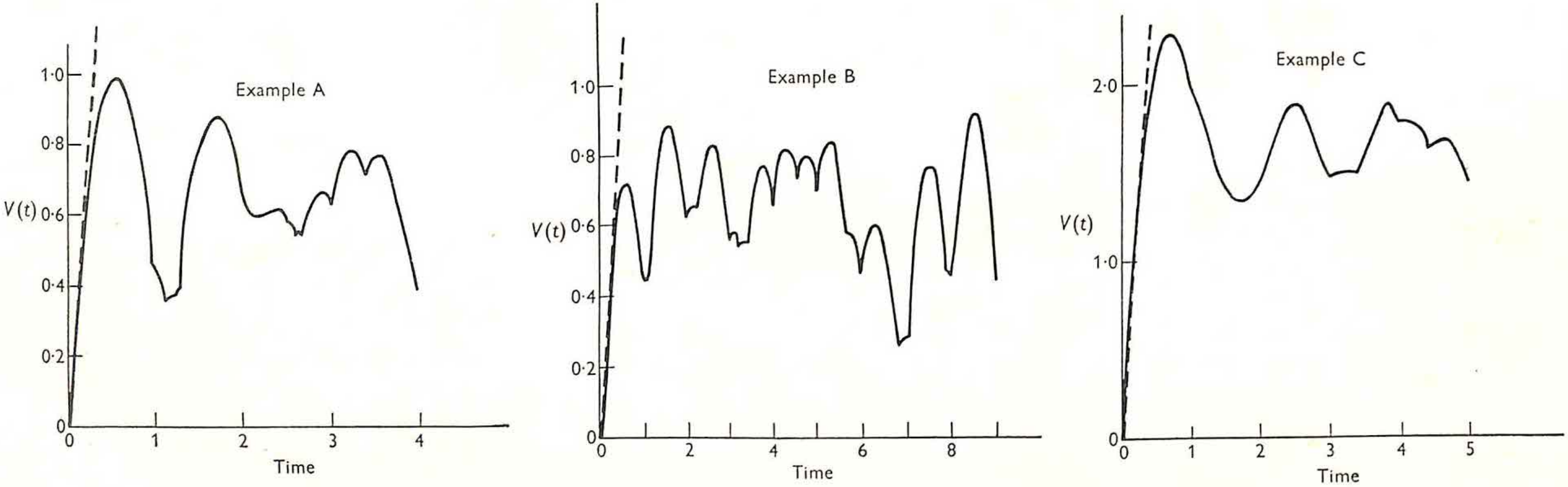


Fig. 4. Variance-time curves. —— pooled periodic sources; ----- random series with same mean interval.

Any method based on the frequency distribution of intervals is clearly insensitive because the frequency curve is very nearly exponential except when N is small. A much better procedure is to use the variance-time curve. To determine this from a set of experimental data, the following method may be used. First divide the series into intervals of length τ such that only a small number of events occur in each interval. Next count the number of events occurring in each interval, so deriving a discrete time series $[x_1, \ldots, x_m]$, where x_i is the number of events in $(i-1)\tau \leqslant t < i\tau$. Add the x's together in blocks of r, giving

$$\left.\begin{aligned} U_1^{(r)} &= x_1 + \ldots + x_r \\ U_2^{(r)} &= x_2 + \ldots + x_{r+1} \\ &\vdots \\ U^{(r}_{m-r+1} &= x_{m-r+1} + \ldots + x_m. \end{aligned}\right\} \tag{28}$$

The $U^{(r)}$'s are the numbers of events occurring in intervals of length $r\tau$ so that an estimate of $V(r\tau)$ may be formed from the corrected sum of squares of the $U^{(r)}$'s. Put $M = m-r+1$ and take as the estimate

$$\hat{V}(r\tau) = \left[\frac{3M}{3M^2 - 3Mr + r^2 - 1}\right] \times \text{corrected sum of squares of } U^{(r)}\text{'s}. \tag{29}$$

Do this for a number of values of r and plot $\hat{V}(r\tau)$ against $r\tau$. The divisor in (29) has been chosen so that if the series is random (i.e. if the x_i are independent Poisson variables with constant mean $\tau\lambda$), then

$$E[\hat{V}(r\tau)] = V(r\tau) = r\tau\lambda, \tag{30}$$

provided that $r < \frac{1}{2}m$.

This analysis is closely connected with the serial correlation analysis of the x_i (Yule, 1945). The variance of $\hat{V}(r\tau)$ can be shown by tedious but elementary methods to be, for a random series,

$$\frac{r\tau\lambda}{3M}(4r^2\tau\lambda + 3r + 2\tau\lambda) + \frac{r\tau\lambda}{3M^2}(-2r^3\tau\lambda + 16r^2\tau\lambda - 24r\tau\lambda + 16\tau\lambda - r^2 + 1) + O\left(\frac{1}{M^3}\right).$$

It was shown in §4 that if the series is the pooled output of N periodic sources, $V(t)$ oscillates about $\frac{1}{6}N$ for large t. This result enables us both to test for the existence of periods and to estimate N. It will be shown in another paper (Cox & Smith, 1953) that if the intervals between the events on any one source have coefficient of variation C, then

$$V(t) \sim C^2\lambda t \quad \text{for large } t, \tag{31}$$

where λ^{-1} is the mean interval between events on the pooled output. Both C and the number of sources can be estimated from the $V(t)$ curve.

Thus our method of using the variance-time curve distinguishes between:

(*a*) a random series;

(*b*) the pooled output of periodic sources;

(*c*) the pooled output of sources on each of which the coefficient of variation of intervals is $C \neq 1$.

The method has been applied to some neurological data. Full details will be given elsewhere.

Summary

Suppose that there are N *sources*, that on the ith source *events* occur at times $\theta_i, 2\theta_i, \ldots$, and that the outputs of the N sources are combined into a single *pooled output*. Statistical properties of the pooled output are investigated. Methods are suggested for distinguishing it from a random series and for estimating N from experimental data. Applications are indicated to experimental psychology and to neuro-physiology.

We are grateful to Dr J. Wishart for helpful comments on the draft of the paper.

REFERENCES

Conrad, R. (1951). *Brit. J. Industr. Med.* **8**, 1.
Cox, D. R. & Smith, W. L. (1953). To be published.
Fatt, P. & Katz, B. (1952). *J. Physiol.* **117**, 109.
Householder, A. S. & Landahl, H. D. (1945). *Mathematical Biophysics of the Central Nervous System*. Bloomington: Principia Press.
Weyl, H. (1916). *Math. Ann.* **77**, 313.
Yule, G. U. (1945). *J. R. Statist. Soc.* **108**, 208.

APPROXIMATE CONFIDENCE INTERVALS

By M. S. BARTLETT
University of Manchester

1. Introduction

It is fairly generally known that asymptotic confidence intervals for an unknown parameter θ may be obtained directly from the likelihood (log) derivative $\partial L/\partial\theta$, where $p\equiv\exp L$ is the probability of the sample (see, for example, M. G. Kendall, 1946, §19·10). Such intervals are asymptotically equivalent to those obtained from the maximum-likelihood estimate $\hat{\theta}$, and have the property of providing asymptotically shortest confidence intervals on the average (see Kendall, §19·12). The direct use of $\partial L/\partial\theta$ has the advantage that its mean and variance are (under the usual differentiation conditions, which will be assumed throughout up to any order required) known *exactly* as 0 and

$$I\equiv E\left\{\left(\frac{\partial L}{\partial\theta}\right)^2\right\}=E\left\{-\frac{\partial^2 L}{\partial\theta^2}\right\},$$

in contrast with the asymptotic properties of $\hat{\theta}$, so that the only approximation introduced as far as the sampling distribution of $\partial L/\partial\theta$ is concerned is in regarding it as normal. This suggests that its use to provide confidence intervals should be approximately valid under fairly wide conditions, and, in so far as we can study the precise distribution of $\partial L/\partial\theta$, may even become exact. The further requirement for a valid confidence interval obtained from any random quantity $T(\theta)$ is that $T(\theta)$ bears a monotonic relation with θ for all samples, yielding a unique and admissible value θ_0 for each critical value T_0.

Now with regard to the normality approximation it is quite feasible to investigate also the higher moments of $\partial L/\partial\theta$, so that a correction to allow, say, for its skewness is not unduly complicated. It is the purpose of this note to develop such a further approximation procedure. It would appear difficult to specify precise and useful conditions under which such an approximation, leading to the use of a quantity $T(\theta)$, say, is necessarily free from possible violation of the further requirements on the relation of $T(\theta)$ with θ, but one or two incidental comments on this point may be helpful. We may expect any correcting terms to have a relatively minor effect except perhaps for very small samples, and thus not to disturb any monotonic relation with θ in the neighbourhood of the critical values of θ. With regard to the intervals which would be obtained from $\partial L/\partial\theta$ if its *exact* distribution were made use of, it is noted that if a sufficient statistic for θ exists the procedure is equivalent to making use of it,* but this does not necessarily imply that an exact confidence interval is possible even in such cases. For example, in small samples a confidence limit for θ, even when the required monotonic relation with θ is present, may yield a value outside the admissible range (this can happen also with the maximum likelihood estimate). In such cases the fact that we cannot always make full use of the sampling distribution means that our admissible confidence limits *over all samples* should have a confidence coefficient *not less than* the level claimed. It might also be remarked that the approximation procedure is sometimes convenient even if a sufficient statistic does exist, in cases where the exact distribution of the sufficient statistic is difficult to handle (e.g. if the observations are from truncated distributions).

---

* This naturally raises the question of how far $\partial L/\partial\theta$ may be regarded in general as a 'theoretical sufficient statistic' for θ (cf. Bartlett, 1952, §6). However, this question is not altogether an unambiguous one, and I think that the answer is strictly only 'yes' if $\partial L/\partial\theta$, as a random variable, is considered *simultaneously* for all possible θ, and is 'no' if it is merely considered for the *single* true value of θ.

Only the case of *one* unknown parameter θ will be considered here, but it is hoped to discuss the case of more than one unknown in a later paper. The extension to confidence *regions* is in principle straightforward, but the case of one unknown plus 'nuisance parameters' is more subtle, being much more dependent on the nature of the sufficiency properties of $\partial L/\partial\theta$ (see earlier footnote; cf. also Bartlett, 1936, §6).

2. MOMENTS OF $\partial L/\partial\theta$

For reference a convenient method will be given of expressing the moments of $\partial L/\partial\theta$ in their simplest terms for evaluation, as an extension of the familiar result

$$E\left\{\left(\frac{\partial L}{\partial\theta}\right)^2\right\} = E\left\{-\frac{\partial^2 L}{\partial\theta^2}\right\}.$$

It is desirable to introduce a more convenient notation, and we shall write

$$E\left\{\frac{\partial L}{\partial\theta}\right\} \equiv L_1, \qquad E\left\{\frac{\partial^2 L}{\partial\theta^2}\right\} \equiv L_2,$$
$$\frac{\partial L_1}{\partial\theta} \equiv {}_1L_1, \qquad \frac{\partial^2 L_1}{\partial\theta^2} \equiv {}_2L_1,$$
$$E\left\{\left(\frac{\partial L}{\partial\theta}\right)^2\right\} \equiv L_1^{(2)}, \quad E\left\{\frac{\partial L}{\partial\theta}\frac{\partial^2 L}{\partial\theta^2}\right\} \equiv (L_1 L_2),$$

etc. Now

$$1 = \int p(\theta+\tau) = \int p(\theta)\exp\left\{\tau\frac{\partial L}{\partial\theta} + \tfrac{1}{2}\tau^2\frac{\partial^2 L}{\partial\theta^2} + \dots\right\},$$

where $\int p(\theta)$ is a formal notation for the probability Stieltjes integral over all possible samples. Expanding the last expression in powers of τ, and equating coefficients, we obtain an infinite series of relations of which the first four are

$$\left.\begin{aligned} &L_1 = 0,\\ &L_2 + L_1^{(2)} = 0,\\ &L_3 + 3(L_1 L_2) + L_1^{(3)} = 0,\\ &L_4 + 3L_2^{(2)} + 4(L_1 L_3) + 6(L_1^{(2)} L_2) + L_1^{(4)} = 0.\end{aligned}\right\} \qquad (1)$$

The first two are the familiar ones already quoted, and the next two involve the third and fourth moments $L_1^{(3)}$ and $L_1^{(4)}$. These are not yet in their simplest terms; thus it is possible to express $L_1^{(3)}$ in 'linear' terms alone. We may, however, obtain an unlimited number of further relations by differentiating the relations (1). We have by differentiating the second and third relations once, noting that

$${}_1L_1^{(2)} \equiv \frac{\partial}{\partial\theta}\int p(\theta)\left(\frac{\partial L}{\partial\theta}\right)^2 = L_1^{(3)} + 2(L_1 L_2),$$

etc.,

$${}_1L_2 + L_1^{(3)} + 2(L_1 L_2) = 0,$$
$${}_1L_3 + 3\{(L_1^{(2)} L_2) + L_2^{(2)} + (L_1 L_3)\} + \{L_1^{(4)} + 3(L_1^{(2)} L_2)\} = 0.$$

Differentiating the former of these two derived relations again, we obtain further

$${}_2L_2+\{L_1^{(4)}+3(L_1^{(2)}L_2)\}+2\{(L_1^{(2)}L_2)+L_2^{(2)}+(L_1\,L_3)\}=0.$$

We find finally, reverting to the full notation for clarity,

$$\kappa_3\left(\frac{\partial L}{\partial\theta}\right)\equiv E\left\{\left(\frac{\partial L}{\partial\theta}\right)^3\right\}=3\frac{\partial I}{\partial\theta}+2E\left\{\frac{\partial^3 L}{\partial\theta^3}\right\},\tag{2}$$

$$\begin{aligned}\kappa_4\left(\frac{\partial L}{\partial\theta}\right)&\equiv E\left\{\left(\frac{\partial L}{\partial\theta}\right)^4\right\}-3I^2\\&=6\frac{\partial^2 I}{\partial\theta^2}+8\frac{\partial}{\partial\theta}E\left\{\frac{\partial^3 L}{\partial\theta^3}\right\}-3E\left\{\frac{\partial^4 L}{\partial\theta^4}\right\}+3\sigma^2\left\{\frac{\partial^2 L}{\partial\theta^2}\right\},\end{aligned}\tag{3}$$

where it will be noticed that κ_4 involves one 'non-linear' quantity, the last variance term involving the mean square of $\partial^2L/\partial\theta^2$. Whether these formulae (2) and (3) are useful will of course depend on the form of L; for example, in the case of the Cauchy distribution it is simpler to evaluate κ_3 and κ_4 directly.

3. APPROXIMATE CORRECTION FOR THE SKEWNESS OF $\partial L/\partial\theta$

It seems most useful to consider polynomial transformations of $\partial L/\partial\theta$ as in principle such transformations may be chosen with the aid of the above relations to annihilate the skewness (and, if necessary, any higher non-zero cumulants) exactly, although it is perhaps unlikely in practice that we shall wish to go further than a first further approximation involving κ_3. Consider the theoretical statistic

$$T_\lambda(\theta)\equiv\frac{\partial L}{\partial\theta}+\lambda\left[\left(\frac{\partial L}{\partial\theta}\right)^2-I\right],\tag{4}$$

which still has zero mean, and variance

$$I+2\lambda\kappa_3+\lambda^2\sigma^2\left\{\left(\frac{\partial L}{\partial\theta}\right)^2\right\}.\tag{5}$$

Its skewness is

$$\kappa_3+3\lambda\sigma^2\left\{\left(\frac{\partial L}{\partial\theta}\right)^2\right\}+O(\lambda^2).\tag{6}$$

As $\sigma^2\{(\partial L/\partial\theta)^2\}=\kappa_4+2I^2$, we therefore choose, to the first order of approximation,

$$\lambda=-\frac{1}{6}\frac{\kappa_3}{I^2},\tag{7}$$

whence

$$T(\theta)\equiv\frac{\partial L}{\partial\theta}-\frac{1}{6}\frac{\kappa_3}{I^2}\left[\left(\frac{\partial L}{\partial\theta}\right)^2-I\right].\tag{8}$$

The variance of T, if we neglect κ_3^2, is still I. More accurate confidence intervals should therefore be obtained if we write

$$T(\theta)=\pm\mu\sqrt{I}\,(\theta),\tag{9}$$

where μ is the appropriate normal deviate for any required significance level, and solve for θ.

Of course if $\kappa_3=0$, the standard approximation remains unaltered to this order. In this particular case, i.e. *when* $\kappa_3=0$, the correction for the next cumulant κ_4 is also recorded. We write

$$T_\nu(\theta)=\frac{\partial L}{\partial\theta}+\nu\left(\frac{\partial L}{\partial\theta}\right)^3\tag{10}$$

with mean zero, variance
$$\sigma^2 = I + 2\nu[\kappa_4 + 3I^2] + O(\nu^2), \tag{11}$$

skewness
$$E\{T_\nu^3\} = 3\nu E\left\{\left(\frac{\partial L}{\partial \theta}\right)^5\right\} + O(\nu^2)$$
$$= 0 + O(\nu^2, \nu\kappa_5), \tag{12}$$

and kurtosis
$$E\{T_\nu^4\} - 3\sigma^4 = \kappa_4 + 3I^2 + 4\nu E\left\{\left(\frac{\partial L}{\partial \theta}\right)^6\right\} - 3\sigma^4 + O(\nu^2),$$

or, after reduction,
$$\kappa_4 + 24\nu I^3 + O(\nu^2, \nu\kappa_4, \nu\kappa_6). \tag{13}$$

Hence we choose
$$\nu = -\frac{1}{24}\frac{\kappa_4}{I^3} \tag{14}$$

and
$$T(\theta) \equiv \frac{\partial L}{\partial \theta} - \frac{1}{24}\frac{\kappa_4}{I^3}\left(\frac{\partial L}{\partial \theta}\right)^3, \tag{15}$$

with variance
$$\sigma_T^2 \sim I(1 - \tfrac{1}{4}\kappa_4/I^2). \tag{16}$$

4. EXAMPLE I

To test out this procedure, we shall try it out on a standard problem whose exact solution is well known, namely, confidence limits for a (normal) sample variance. This example is taken because the sample variance s^2 is quite skew for a moderate number of degrees of freedom, and there is some interest in seeing how far the method can cope with this skewness. We have

$$\frac{\partial L}{\partial \theta} = \frac{n}{2\theta^2}(s^2 - \theta),$$

$$-\frac{\partial^2 L}{\partial \theta^2} = -\frac{n}{2\theta^2} + \frac{ns^2}{\theta^3}, \qquad I = \frac{n}{2\theta^2},$$

$$\frac{\partial^3 L}{\partial \theta^3} = -\frac{n}{\theta^3} + \frac{3ns^2}{\theta^4}, \quad E\left\{\frac{\partial^3 L}{\partial \theta^3}\right\} = \frac{2n}{\theta^3},$$

$$\frac{\partial I}{\partial \theta} = -\frac{n}{\theta^3}, \qquad \kappa_3 = \frac{n}{\theta^3},$$

the value of κ_3 checking of course with the value inferred from the properties of s^2. The standard approximation for the confidence interval would amount to solving

$$\frac{n}{2\theta^2}(1 - \theta) = \pm\mu\sqrt{\left(\frac{n}{2\theta^2}\right)}, \tag{17}$$

where for convenience we have expressed confidence limits θ in terms of s^2, i.e. have put $s^2 = 1$. The more accurate approximation suggested replaces (17) by the equation

$$\frac{n}{2\theta^2}(1 - \theta) - \frac{1}{6}\frac{n}{\theta^3}\left[(1-\theta)^2 - \frac{2\theta^2}{n}\right] = \pm\mu\sqrt{\left(\frac{n}{2\theta^2}\right)}. \tag{18}$$

This as it stands no longer yields a simple quadratic equation for θ, but to the same order of approximation we may replace the square bracket by its value from (17), viz. $2\theta^2(\mu^2 - 1)/n$, and (18) then also reduces to a quadratic (i.e. to a linear equation for each sign of μ). A corresponding simplification is available in general, for the crude interval is obtained from

$$\frac{\partial L}{\partial \theta} = \pm \mu \sqrt{I}, \tag{19}$$

and substituting in $T(\theta)$ in (9), we obtain

$$\frac{\partial L}{\partial \theta} - \frac{1}{6}\frac{\kappa_3}{I}(\mu^2 - 1) = \pm \mu \sqrt{I} \tag{20}$$

as the corresponding modification of the general equation (9).†

In arithmetical work where $\partial L/\partial \theta$ is calculated for various values of θ, either (20) or (9) may be used. The original approximation (19) is most conveniently plotted as

$$\frac{\partial L}{\partial \theta}\bigg/\sqrt{I} = \pm \mu, \tag{21}$$

for the values of θ for which the left-hand side passes any chosen value of μ can be read off from the same graph. Equation (9) similarly is written

$$\frac{\partial L}{\partial \theta}\bigg/\sqrt{I} - \frac{1}{6}\frac{\kappa_3}{I^{\frac{3}{2}}}\left[\left(\frac{\partial L}{\partial \theta}\bigg/\sqrt{I}\right)^2 - 1\right] = \pm \mu, \tag{22}$$

and may be used in the same way. The equation equivalent to (20), viz.

$$\frac{\partial L}{\partial \theta}\bigg/\sqrt{I} - \frac{1}{6}\frac{\kappa_3}{I^{\frac{3}{2}}}[\mu^2 - 1] = \pm \mu, \tag{23}$$

may only be used at the one chosen value of μ, but has the slight advantage that if interpolation for θ (or some suitable function $g(\theta)$) is approximately linear in (21), it is likely to be so also for (23), rather than for (22).

Table 1

n	Exact values				First or standard approximation				Second or further approximation			
	Lower 0·01	Lower 0·05	Upper 0·05	Upper 0·01	Lower 0·01	Lower 0·05	Upper 0·05	Upper 0·01	Lower 0·01	Lower 0·05	Upper 0·05	Upper 0·01
5	0·331	0·452	4·37	9·03	0·405	0·490	∞*	∞*	0·327	0·441	5·35	8·55
10	0·431	0·546	2·54	3·91	0·490	0·576	3·78	∞*	0·428	0·541	2·65	3·94
20	0·532	0·637	1·84	2·42	0·576	0·658	2·08	3·78	0·531	0·634	1·86	2·43
30	0·589	0·685	1·62	2·01	0·625	0·702	1·74	2·50	0·589	0·684	1·63	2·01

* In this example the approximations break down at the upper limit for small n and P; this occurs with the first approximation even for some of the values in the table.

To return to (17) and (18), the resulting limits are compared in Table 1 with the correct limits obtained from the distribution of s^2, for upper and lower significance limits $P = 0{\cdot}05$ and 0·01, or total significance limits of 0·10 and 0·02, and representative values 5, 10, 20, and 30 for n.

---

† This further approximation has been used when obtaining the values given in Tables 1 and 2.

It will be seen that even the second approximation is not too good at the upper limit for $n = 5$, and would be inferior to at least one other available approximation for the significance limits for s^2, but the important point to remember is that it is a much more general method, and this example suggests that it should be a considerable improvement over the first approximation, which is still hardly satisfactory in this example for n as large as 30.

5. EXAMPLE II

As a second example we shall examine a standard discrete distribution, choosing for simplicity the Poisson rather than the binomial. Here we easily find, for θ the unknown mean and x an observation,

$$\frac{\partial L}{\partial \theta} = \frac{x}{\theta} - 1, \quad I = \frac{1}{\theta},$$

$$E\left\{\frac{\partial^3 L}{\partial \theta^3}\right\} = \frac{2}{\theta^2}, \qquad \kappa_3 = \frac{1}{\theta^2},$$

whence the crude approximation gives the equation

$$\frac{x}{\theta} - 1 = \pm \frac{\mu}{\sqrt{\theta}}, \tag{24}$$

and the further approximation (corresponding to (20))

$$\frac{x}{\theta} - 1 - \frac{1}{6\theta}(\mu^2 - 1) = \pm \frac{\mu}{\sqrt{\theta}}. \tag{25}$$

Of course for discrete distributions a confidence interval (at least if obtained from the data directly without introduction of extra randomization devices) can only be given as an inequality for the confidence level. This is perhaps hardly crucial if an approximate method

Table 2

x	Exact values*				First or standard approximation				Second or further approximation			
	Lower 0·01	Lower 0·05	Upper 0·05	Upper 0·01	Lower 0·01	Lower 0·05	Upper 0·05	Upper 0·01	Lower 0·01	Lower 0·05	Upper 0·05	Upper 0·01
0	0	0	3·00	4·61	†	†	3·64	6·37	†	†	3·12	4·93
1	0·01	0·05	4·74	6·64	0·04	0·07	5·28	8·14	†	0·02	4·83	6·86
2	0·15	0·36	6·30	8·41	0·28	0·43	6·79	9·77	0·09	0·31	6·37	8·58
3	0·44	0·82	7·75	10·05	0·64	0·92	8·22	11·33	0·36	0·77	7·81	10·19
5	1·28	1·97	10·51	13·11	1·58	2·11	10·94	14·30	1·21	1·93	10·56	13·22
10	4·13	5·43	16·96	20·14	4·54	5·61	17·35	21·21	4·07	5·40	17·00	20·23
20	11·08	13·25	29·06	33·10	11·58	13·46	29·42	34·08	11·04	13·23	29·09	33·16
30	18·74	21·59	40·69	45·40	19·29	21·82	41·04	46·33	18·71	21·58	40·71	45·45

* Quoted from Garwood (1936).

† The first approximation gives $\theta = 0$ here only if the continuity correction is *not* used. The second approximation no longer gives near $\theta = 0$ a monotonic relation between $T(\theta)$ and θ for small x, and thus breaks down near $\theta = 0$. (The inadequacy of the second approximation for very small θ is hardly unexpected, for the Poisson mean θ in this example effectively takes the place of n, the size of the sample.)

is in any case being used, but it reminds us that a still further approximation is now involved in obtaining limits from normal theory. For a purely discrete distribution as in this example some gain would be expected from the introduction of the usual continuity correction, which implies using $x-\frac{1}{2}$ when obtaining the lower limit for θ and $x+\frac{1}{2}$ for the upper limit. This correction has therefore been used throughout, though from an examination of one or two values the consequent gain for the first crude approximation appeared much more doubtful than for the second.

It will be seen from Table 2 that the second approximation is a considerable improvement over the first (except near $\theta = 0$), and is in fact somewhat better for small x at the upper limit than we might have anticipated from the experience with the first example. The exact values quoted were readily obtained by Garwood from χ^2 significance limits, and, as in Example I, it must be emphasized that the purpose of examining the present approximation in these standard examples is merely to obtain an idea of its accuracy.

6. EXAMPLE III

The problem which actually initiated this inquiry is being discussed in detail elsewhere, but it may be helpful to indicate it here. From the tracks of certain cosmic ray particles which 'decay' into other fundamental particles, it was required to estimate the 'decay' parameter θ in the lifetime distribution

$$f(t)\,dt = e^{-t/\theta}\,dt/\theta. \tag{26}$$

This is simple enough in the case of unlimited track length, but for tracks in a chamber of finite size, not all particles will decay; the chance of doing so moreover varies with each particle because its time of passage through the chamber depends on its momentum. The situation is not always the same as this and is often complicated by further observational difficulties, but for simplicity here we shall suppose we have just N particles of one type, with effective times T_s $(s = 1 \dots N)$ in the chamber, of which n decay at times t_r $(r = 1 \dots n)$. The likelihood function in this case is a mixture of finite probability factors and densities, but this causes no difficulty. For such data we have, as the probability of the sth undecayed particle is $Q_s = e^{-T_s/\theta}$, and the probability density of the rth decayed particle is $f(t_r) = e^{-t_r/\theta}/\theta$,

$$\frac{\partial L}{\partial \theta} = -\frac{n}{\theta} + \sum_{r=1}^{n} \frac{t_r}{\theta^2} + \sum_{s=n+1}^{N} \frac{T_s}{\theta^2}, \tag{27}$$

giving in this particular case the immediate estimate

$$\hat{\theta} = \frac{1}{n}\left(\sum_{r=1}^{n} t_r + \sum_{s=n+1}^{N} T_s\right). \tag{28}$$

(The record of undecayed particles is not always available; in that case the estimation equation is different and not quite such a simple solution $\hat{\theta}$ is possible from the combined data.) It is not difficult to show that for equation (27)

$$I = \sum_{s=1}^{N} \frac{P_s}{\theta^2}, \tag{29}$$

where $P_s = 1 - Q_s$. Thus the first approximation for the confidence interval would be obtained from equation (19), with $\partial L/\partial \theta$ given now by (27) and I by (29). In the limiting case when all T_s become large, the problem reduces to Example I, for each measurement t from the exponential distribution (26) is equivalent to a χ^2 with two degrees of freedom, and the mean t would be equivalent to a χ^2 (or s^2) with $2N$ degrees of freedom. In practice the T_s are often small, but in any case, in any attempt to obtain a confidence interval for θ, it would seem advisable when possible to go to the further approximation (9) or (20).

It is not too difficult, at least in the case covered by equation (27), to obtain the exact moment-generating function $M(\phi)$ of $\theta^2 \partial L/\partial\theta$, remembering that the chance of decay somewhere in the chamber for each particle is $P_s\,(s = 1 \ldots N)$. The answer comes out to be $M(\phi) = \prod_{s=1}^{N} M_s(\phi)$, with

$$M_s(\phi) = e^{-\phi\theta}[1 - Q_s e^{\phi T_s}\{1-(1-\theta\phi)\,e^{\theta\phi}\}]/(1-\theta\phi). \tag{30}$$

Expanding $\log M(\phi)$, we find for the cumulants of $\theta^2 \partial L/\partial\theta$

$$\left.\begin{aligned} \kappa_1 &= 0, \quad \kappa_2 = \theta^2 \sum_{s=1}^{N} P_s, \\ \kappa_3 &= 2\theta^3 \sum_{s=1}^{N} P_s - 3\theta^2 \sum_{s=1}^{N} Q_s T_s, \\ \kappa_4 &= 6N\theta^4 - 8\theta^3 \sum_{s=1}^{N} Q_s T_s - 6\theta^2 \sum_{s=1}^{N} T_s^2 Q_s + \theta^4 \sum_{s=1}^{N} Q_s - 3\theta^4 \sum_{s=1}^{N} Q_s^2. \end{aligned}\right\} \tag{31}$$

However, as the distribution of $\partial L/\partial\theta$ is a mixture of continuous and discrete components (for there is a finite chance that none of the N particles decays), it would be difficult to make use of the exact distribution of $\partial L/\partial\theta$ in this particular case, and the use of the cumulants would be equivalent to using the general method developed earlier in this paper. The formulae in (31), while first obtained directly by the above method, agree of course with the results derived via the general method, and the expressions for κ_3 and κ_4 were checked by this means.

For maximum-likelihood or confidence interval equations not directly soluble, iterative or interpolative methods may usually be used (see equations (21), (22) and (23), § 4).

I am indebted to Mrs A. Linnert for assistance with the calculations for Tables 1 and 2.

REFERENCES

BARTLETT, M. S. (1936). Statistical information and properties of sufficiency. *Proc. Roy. Soc.* A, **154**, 124.

BARTLETT, M. S. (1952). The statistical significance of odd bits of information. *Biometrika*, **39**, 228.

GARWOOD, F. (1936). Fiducial limits for the Poisson distribution. *Biometrika*, **28**, 437.

KENDALL, M. G. (1946). *The Advanced Theory of Statistics*, 1st ed. **2**. London: Griffin.

Notes added in proof. (i) The coefficients in the polynomial transformations considered in § 3 are equivalent to those in a series expansion given by E. A. Cornish & R. A. Fisher in § 8 of their paper 'Moments and cumulants in the specification of distributions,' *Rev. Inst. Int. Statis.* (1937), **4**, which should be consulted if further terms of the expansion are required.

(ii) The problem referred to in § 6 I have discussed further in my paper 'On the statistical estimation of mean life-time,' *Phil. Mag.* (7th series), 1953, **44** (in the Press).

INCOMPLETE AND ABSOLUTE MOMENTS OF THE MULTIVARIATE NORMAL DISTRIBUTION WITH SOME APPLICATIONS

By A. R. KAMAT,
University College, London

1. Introduction

It is a well-known property of normally distributed error variables $x_1, x_2, x_3, \ldots$ that the process of totalling or averaging the error produces a normal distribution. This property also, of course, holds if the errors are correlated 'normally'. Occasions arise, however, when it is necessary to consider the total numerical error $T = |x_1| + |x_2| + |x_3| + \ldots$ (or its average). Tricomi (1936) has investigated the statistic T as a measure of dispersion when the variables are uncorrelated and are of equal variance. Arley & Hald (1950) have proposed the statistic $\sum_{i=1}^{n-1} |x_i - x_{i+1}|/(n-1)$ as a measure of dispersion.*

In certain technological applications also it is found that individual errors compound into a total error obtained by the addition of numerical values of individual errors. We may mention the following example: In the manufacture of metal rods it is found that the lengths of individual rods differ from specification by a normal error variable. Pairs of rods with lengths x_1 and x_2 are assembled in a manner requiring the longer of the pair to be filed down to the length of the shorter one. If the cross-section of the rods is constant, the total loss of weight through filing is proportional to $\Sigma |x_1 - x_2|$, which is a sum of independent mod-normal variates. Occasions also arise when special mechanisms give rise to a total error which is a sum of mod-normal errors or a sum of normal and mod-normal errors, although in such situations there are only two or three error components. For instance, in the manufacture of screw gauges two main errors arise, an error x in the diameter of the gauge and an error y in its pitch. These two result in a total vertical displacement of the points of contact of screw and gauge which is proportional to $x + c|y|$, where c depends on the pitch angle.

While dealing with some special cases of correlated error variables, the present author was concerned with the evaluation of the moments of T and was thereby naturally led to the calculation of the 'absolute moments' of the multivariate normal distribution, namely,

$$(l, m, n, \ldots) = \int_{-\infty}^{\infty} \cdots \int |x_1|^l \, |x_2|^m \, |x_3|^n \ldots p(x)\, dx, \tag{1}$$

where $p(x)$ is the probability-density function of $x_1, x_2, \ldots, x_n$, these variates having zero means and given variances and covariances. Although we shall not be able to deal with them in this paper, certain other statistics require for their study the knowledge of the contribution to these absolute moments *separately from each quadrant*, i.e. integrals of the type

$$[l, m, n, \ldots] = \int_0^{\infty} \cdots \int x_1^l x_2^m x_3^n \ldots p(x)\, dx. \tag{2}$$

---

* A detailed investigation of this statistic is the subject of a further paper by the present author printed on pp. 116–27 below.

It is clear that the contributions from different quadrants can be obtained from (2) by changing the signs of the appropriate elements of the correlation matrix and that (1) can be found as the sum of all these contributions.

In §2 of the paper we shall deal with the evaluation of $[l, m, n]$ for the bivariate and the trivariate cases and hence with that of (l, m, n). Since the work was put in hand, papers have been published by Nabeya (1951, 1952) giving the absolute moments for the bivariate and the trivariate cases and a generalization in the form of a multiple integral for the multivariate case.* Nabeya's method is not, however, applicable to the evaluation of $[l, m, n]$, and the methods developed in the present paper are also believed to be simpler for evaluating certain of the absolute moments (l, m, n). In §3, $[l, m, n, \ldots]$ and $(l, m, n, \ldots)$ are obtained for a multivariate distribution in the form of power series of correlation coefficients. Finally, §4 is concerned with the application of the results to the distribution of T and related statistics.

2. DERIVATION OF THE ABSOLUTE MOMENTS

In the following, without loss of generality, the variates are measured in terms of their standard deviations, so that the n-variate normal distribution is given by

$$p(x_1, x_2, \ldots, x_n) = (2\pi)^{-\frac{1}{2}n}\, \omega^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\omega}\Sigma\omega_{rs}x_r x_s\right\} \quad (r, s = 1, 2, \ldots, n),$$

where ω is the determinant of the correlation matrix, ω_{rs} are its cofactors, and the quadratic form $\Sigma\omega_{rs}x_r x_s$ is positive definite.

2·1. *Univariate case*

For the univariate case the results are well known and are easily found by direct integration:

$$[r] = \tfrac{1}{2}(r) = (2\pi)^{-\frac{1}{2}}\int_0^\infty x^r\, e^{-\frac{1}{2}x^2} dx = \pi^{-\frac{1}{2}} 2^{\frac{1}{2}(r-2)}\, \Gamma\left(\frac{r+1}{2}\right). \tag{3}$$

The following representation is given only to illustrate the method which will be followed in the bivariate and trivariate cases. Defining

$$R_0 = \int_0^\infty e^{-\alpha z^2} dz = \frac{1}{2}\sqrt{\frac{\pi}{\alpha}}, \quad R_1 = \int_0^\infty z\, e^{-\alpha z^2} dz = \frac{1}{2\alpha} \quad (\alpha > 0) \tag{4}$$

it is easy to see that

$$[r] = \tfrac{1}{2}(r) = \pi^{-\frac{1}{2}} 2^{\frac{1}{2}r} P_r, \tag{5}$$

where

$$P_{2k} = \left[(-1)^k \frac{\partial^k P_0}{\partial \alpha^k}\right]_{\alpha=1}, \quad P_{2k+1} = \left[(-1)^k \frac{\partial^k R_1}{\partial \alpha^k}\right]_{\alpha=1}, \tag{6}$$

so that the familiar results (3) are obtained.

2·2. *Bivariate case*

For the bivariate case

$$[m, n] = \iint_0^\infty x^m y^n p(x, y)\, dx\, dy, \quad (m, n) = \iint_{-\infty}^{\infty} |\, x^m \,|\, |\, y^n \,|\, p(x, y)\, dx\, dy, \tag{7}$$

where

$$p(x, y) = (2\pi)^{-1}(1-\rho^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(x^2 + y^2 - 2\rho xy)\right\},$$

---

* We gratefully acknowledge the courtesy of Mr S. Nabeya for letting us see a manuscript of his second paper in advance of its publication.

EDNL. & PSY. LIBRARY
68·70
J·15

the substitution $x' = [2(1-\rho^2)]^{-\frac{1}{2}}x$, $y' = [2(1-\rho^2)]^{-\frac{1}{2}}y$ transforms (7) into

and

$$\left.\begin{aligned}[m,n] &= 2^{-\frac{1}{2}}\pi^{-1}[2(1-\rho^2)]^{\frac{1}{2}(m+n+1)}P_{m,n}\\ (m,n) &= 2^{\frac{1}{2}}\pi^{-1}[2(1-\rho^2)]^{\frac{1}{2}(m+n+1)}(P_{m,n}+Q_{m,n}),\end{aligned}\right\}\tag{8}$$

where

$$P_{m,n} = \iint_0^\infty x^m y^n \exp\{-(x^2+y^2-2\rho xy)\}\,dx\,dy,$$

and $Q_{m,n}$ is obtained from $P_{m,n}$ by changing ρ to $-\rho$. Defining three basic integrals

$$R_{00} = \iint_0^\infty f(x,y)\,dx\,dy,\quad R_{10} = \iint_0^\infty xf(x,y)\,dx\,dy,\quad R_{01} = \iint_0^\infty yf(x,y)\,dx\,dy,\tag{9}$$

where $f(x,y) = \exp\{-(\omega_{11}x^2+\omega_{22}y^2+2\omega_{12}xy)\}$ and the quadratic form $\omega_{11}x^2+\omega_{22}y^2+2\omega_{12}xy$ is positive definite, it is easy to see that $P_{m,n}$ can be obtained as a higher derivative of R_{00} or R_{10} or R_{01} (as the case may be) with respect to the parameters $\omega_{11}, \omega_{22}, \omega_{12}$, with the substitution of $\omega_{11} = \omega_{22} = 1$ and $\omega_{12} = -\rho$ after differentiation. $Q_{m,n}$ is obtained from $P_{m,n}$ by simply replacing ρ by $-\rho$. For instance,

$$P_{3,3} = \left[(-\tfrac{1}{2})^3\frac{\partial^3 R_{00}}{\partial\omega_{12}^3}\right]_{\omega_{11}=\omega_{22}=1,\,\omega_{12}=-\rho}$$

and

$$P_{3,4} = \left[(-1)^3\frac{\partial^3 R_{10}}{\partial\omega_{11}\,\partial\omega_{22}^2}\right]_{\omega_{11}=\omega_{22}=1,\,\omega_{12}=-\rho}.$$

2·3. *Bivariate case: Evaluation of the basic integrals*

It now remains to evaluate the basic integrals R_{00}, R_{10} and R_{01}. Changing to polar co-ordinates and using standard integration, we obtain for R_{00} the two expressions which will be used in the following:

$$\left.\begin{aligned}R_{00} &= \tfrac{1}{2}(\omega_{11}\omega_{22}-\omega_{12}^2)^{-\frac{1}{2}}\left[\tfrac{1}{2}\pi-\sin^{-1}(\omega_{12}(\omega_{11}\omega_{22})^{-\frac{1}{2}})\right]\\ &= \tfrac{1}{2}(\omega_{11}\omega_{22}-\omega_{12}^2)^{-\frac{1}{2}}\cos^{-1}(\omega_{12}(\omega_{11}\omega_{22})^{-\frac{1}{2}}).\end{aligned}\right\}\tag{10}$$

The integrals R_{10}, R_{01} may be evaluated in the same way or by the following method which will also be used in the trivariate case. Clearly the transformation $\sqrt{\omega_{11}}\,x = x'$, $\sqrt{\omega_{22}}\,y = y'$ transforms

$$R_{10} = \omega_{11}^{-1}\omega_{22}^{-\frac{1}{2}}J_x,\tag{11}$$

where

$$J_x = \iint_0^\infty x\exp\{-(x^2+y^2+2axy)\}\,dx\,dy,\quad a = (\omega_{11}\omega_{22})^{-\frac{1}{2}}\omega_{12}.\tag{12}$$

Similarly,

$$R_{01} = \omega_{11}^{-\frac{1}{2}}\omega_{22}^{-1}J_y,\tag{13}$$

where

$$J_y = \iint_0^\infty y\exp\{-(x^2+y^2+2axy)\}\,dx\,dy.\tag{14}$$

Now

$$J_x + aJ_y = \iint_0^\infty (x+ay)\exp\{-(x^2+y^2+2axy)\}\,dx\,dy,$$

and introducing $x' = x+ay$, $y' = y$ we have

$$J_x + aJ_y = \tfrac{1}{2}\int_0^\infty e^{-y'^2}\,dy' = \tfrac{1}{4}\sqrt{\pi}.\tag{15}$$

Similarly
$$aJ_x + J_y = \tfrac{1}{4}\sqrt{\pi}. \tag{16}$$

Solving (15) and (16)
$$J_x = J_y = \tfrac{1}{4}\sqrt{\pi}\,(1+a)^{-1}$$

and substituting for a, R_{10} and R_{01} are obtained as

$$\left.\begin{aligned} R_{10} &= \tfrac{1}{4}\sqrt{\pi}\,\omega_{11}^{-\frac{1}{2}}(\omega_{11}^{\frac{1}{2}}\omega_{22}^{\frac{1}{2}}+\omega_{12})^{-1}, \\ R_{01} &= \tfrac{1}{4}\sqrt{\pi}\,\omega_{22}^{-\frac{1}{2}}(\omega_{11}^{\frac{1}{2}}\omega_{22}^{\frac{1}{2}}+\omega_{12})^{-1}. \end{aligned}\right\} \tag{17}$$

2·4. *Trivariate case*

For the trivariate normal distribution the contribution from the positive quadrant is given by

$$[l,m,n] = (2\pi)^{-\frac{3}{2}}\,\omega^{-\frac{1}{2}}\iiint_0^\infty x_1^l x_2^m x_3^n \exp\left\{-\frac{1}{2\omega}\,\Sigma\omega_{rs}x_r x_s\right\} dx_1\,dx_2\,dx_3 \quad (r,s=1,2,3). \tag{18}$$

The transformation $x_r' = (2\omega)^{-\frac{1}{2}}\,x_r$ carries (18) into

$$[l,m,n] = \pi^{-\frac{3}{2}}\,2^{\frac{1}{2}(l+m+n)}\omega^{\frac{1}{2}(l+m+n+2)}P_{l,m,n}, \tag{19}$$

where
$$P_{l,m,n} = \iiint_0^\infty x_1^l x_2^m x_3^n \exp\{-\Sigma\omega_{rs}x_r x_s\}\,dx_1\,dx_2\,dx_3. \tag{20}$$

In the same manner we can show that the absolute moment

$$(l,m,n) = \pi^{-\frac{3}{2}}(2\omega)^{\frac{1}{2}(l+m+n+2)}\,[P^{(1)}_{l,m,n}+P^{(2)}_{l,m,n}+P^{(3)}_{l,m,n}+P^{(4)}_{l,m,n}],$$

where $P^{(i)}_{l,m,n}$ $(i=1,2,3,4)$ are obtained from (20) with the signs of $(\pm\omega_{rs})$, $r \neq s$ in each $P^{(i)}_{l,m,n}$ taken in accordance with the sign-pattern shown below:

i	ω_{12}	ω_{23}	ω_{13}
1	+	+	+
2	+	−	−
3	−	+	−
4	−	−	+

(21)

As in the bivariate case, it can be shown that $P_{l,m,n}$ can be obtained by differentiations with respect to ω_{rs} from the four basic integrals defined as follows:

$$\left.\begin{aligned} R_{000} &= \iiint_0^\infty f(x_1,x_2,x_3)\,dx_1\,dx_2\,dx_3, & R_{100} &= \iiint_0^\infty x_1 f(x_1,x_2,x_3)\,dx_1\,dx_2\,dx_3, \\ R_{010} &= \iiint_0^\infty x_2 f(x_1,x_2,x_3)\,dx_1\,dx_2\,dx_3, & R_{001} &= \iiint_0^\infty x_3 f(x_1,x_2,x_3)\,dx_1\,dx_2\,dx_3, \end{aligned}\right\} \tag{22}$$

where
$$f(x_1,x_2,x_3) = \exp\{-\Sigma\omega_{rs}x_r x_s\} \quad (r,s=1,2,3).$$

The four results $P^{(i)}_{l,m,n}$ $(i=1,2,3,4)$ are obtained essentially by the same differentiations, the difference being merely in the insertion after differentiation of the appropriate signs of $(\pm\omega_{rs})$, $r \neq s$ in accordance with (21).

2·5. *Trivariate case: Evaluation of basic integrals R_0 and R_1*

(*a*) *Evaluation of R_{000}.*

The transformation

$$\sqrt{\omega_{11}}x_1 = x, \quad \sqrt{\omega_{22}}x_2 = y, \quad \sqrt{\omega_{33}}x_3 = z$$

gives

$$R_{000} = (\omega_{11}\omega_{22}\omega_{33})^{-\frac{1}{2}} \iiint\limits_0^\infty \exp\{-(x^2+y^2+z^2+2cxy+2ayz+2bzx)\}\,dx\,dy\,dz, \tag{23}$$

where

$$a = (\omega_{22}\omega_{33})^{-\frac{1}{2}}\,\omega_{23}, \quad b = (\omega_{33}\omega_{11})^{-\frac{1}{2}}\,\omega_{13}, \quad c = (\omega_{11}\omega_{22})^{-\frac{1}{2}}\,\omega_{12}. \tag{24}$$

The transformation $x = r\sin\theta\cos\phi$, $y = r\cos\theta$, $z = r\sin\theta\sin\phi$ transforms the integral on the right-hand side of (23) into

$$\int_{\phi=0}^{\frac{1}{2}\pi}\int_{\theta=0}^{\frac{1}{2}\pi}\int_{r=0}^{\infty} \exp\{-r^2[1+2(a\sin\phi+b\cos\phi)\sin\theta\cos\theta+2c\sin^2\theta\sin\phi\cos\phi]\}\,r^2\sin\theta\,dr\,d\theta\,d\phi$$

$$= \tfrac{1}{4}\sqrt{\pi}\int_0^{\frac{1}{2}\pi} d\phi \int_0^{\frac{1}{2}\pi} [1+2(a\sin\phi+b\cos\phi)\sin\theta\cos\theta+2c\sin^2\theta\sin\phi\cos\phi]^{-\frac{3}{2}}\sin\theta\,d\theta.$$

The substitution $t = \tan\theta$ gives, on integration,

$$R_{000} = \tfrac{1}{4}\sqrt{\pi}\,(\omega_{11}\omega_{22}\omega_{33})^{-\frac{1}{2}} \int_0^{\frac{1}{2}\pi} [BA^{-\frac{1}{2}}(B^2-A)^{-1} - (B^2-A)^{-1}]\,d\phi, \tag{25}$$

where

$$A = 1+2c\sin\phi\cos\phi, \quad B = a\sin\phi+b\cos\phi.$$

The second integral in (25) is evaluated at once; the first is integrated with the substitution $\tan\phi + c = \tan\phi'$. After some simplification, we obtain

$$R_{000} = \tfrac{1}{4}\sqrt{\pi}\,(\omega_{11}\omega_{22}\omega_{33})^{-\frac{1}{2}}\,\Delta^{-\frac{1}{2}}\left\{\tfrac{1}{2}\pi - \sum_{a,\,b,\,c}\tan^{-1}\left(\frac{a-bc}{\Delta^{\frac{1}{2}}}\right)\right\}, \tag{26}$$

where

$$\Delta = \begin{vmatrix} 1 & c & b \\ c & 1 & a \\ b & a & 1 \end{vmatrix} = (\omega_{11}\omega_{22}\omega_{33})^{-1}\,W, \tag{27}$$

W being the determinant

$$\begin{vmatrix} \omega_{11} & \omega_{12} & \omega_{13} \\ \omega_{12} & \omega_{22} & \omega_{23} \\ \omega_{13} & \omega_{23} & \omega_{33} \end{vmatrix}. \tag{28}$$

Since $a - bc = \omega_{11}^{-1}(\omega_{22}\omega_{33})^{-\frac{1}{2}}(\omega_{23}\omega_{11} - \omega_{12}\omega_{13})$, etc., we have finally

$$R_{000} = \tfrac{1}{4}\sqrt{\pi}\,W^{-\frac{1}{2}}\left\{\tfrac{1}{2}\pi + \sum_{1,\,2,\,3}\tan^{-1}\left(\frac{\omega_{12}\omega_{13} - \omega_{23}\omega_{11}}{\sqrt{(W\omega_{11})}}\right)\right\}. \tag{29}$$

(*b*) *Evaluation of R_{100}, R_{010}, R_{001}.*

Employing the same transformation,

$$R_{100} = \omega_{11}^{-1}(\omega_{22}\omega_{33})^{-\frac{1}{2}}\,J_x,$$

where

$$J_x = \iiint\limits_0^\infty x\exp\{-(x^2+y^2+z^2+2cxy+2ayz+2bzx)\}\,dx\,dy\,dz, \tag{30}$$

a, b, c having the same meaning as in (24). Similarly

$$R_{010} = \omega_{22}^{-1}(\omega_{33}\omega_{11})^{-\frac{1}{2}}\,J_y, \tag{31}$$

$$R_{001} = \omega_{33}^{-1}(\omega_{11}\omega_{22})^{-\frac{1}{2}}\,J_z, \tag{32}$$

where J_y, J_z are defined in a manner similar to J_x. The substitution $x' = x+cy+bz$, $y' = y$, $z' = z$ gives

$$J_x+cJ_y+bJ_z = \iiint_0^\infty x' \exp\{-[x'^2+(1-b^2)\,y'^2+(1-c^2)\,z'^2+2(a-bc)\,y'z']\}\,dx'\,dy'\,dz'$$

$$= \frac{1}{2}\iint_0^\infty \exp\{-(y'^2+z'^2+2ay'z')\}\,dy'\,dz' = I_a, \tag{33}$$

where, from (10),
$$I_a = \tfrac{1}{4}(1-a^2)^{-\frac{1}{2}}\cos^{-1}(a). \tag{34}$$

Similarly,
$$cJ_x+J_y+aJ_z = I_b, \tag{35}$$

$$bJ_x+aJ_y+J_z = I_c, \tag{36}$$

where I_b, I_c are defined as in (27). Solving these linear equations (33), (35) and (36) we have

$$J_x = \Delta^{-1}\{(1-a^2)\,I_a+(ab-c)\,I_b+(ac-b)\,I_c\}, \tag{37}$$

where Δ is defined as in (34) above. Expressing a, b, c in terms of ω_{rs}, we have

$$R_{100} = \tfrac{1}{4}W^{-1}\Big[(\omega_{22}\omega_{33}-\omega_{23}^2)^{\frac{1}{2}}\cos^{-1}(\omega_{23}(\omega_{22}\omega_{33})^{-\frac{1}{2}})$$
$$+\frac{\omega_{23}\omega_{13}-\omega_{12}\omega_{33}}{(\omega_{33}\omega_{11}-\omega_{13}^2)^{\frac{1}{2}}}\cos^{-1}(\omega_{13}(\omega_{11}\omega_{33})^{-\frac{1}{2}})+\frac{\omega_{23}\omega_{12}-\omega_{13}\omega_{22}}{(\omega_{11}\omega_{22}-\omega_{12}^2)^{\frac{1}{2}}}\cos^{-1}(\omega_{12}(\omega_{11}\omega_{22})^{-\frac{1}{2}})\Big], \tag{38}$$

where W is defined as in (28). R_{010}, R_{001} will have similar expressions.

2·6. *Table of results*

We give below a list of $[m, n]$ and (m, n) for the bivariate case and (l, m, n) for the trivariate case up to weight six. For a fuller list of (l, m, n) the reader is referred to Nabeya (1951, 1952). Values of $[l, m, n]$ for the trivariate case have not been given here because of their lengthy expressions. The following observations may be made: (1) All $[m, n]$ and some $[l, m, n]$ can be obtained by direct integration, but the procedure involved is very tedious. (2) When the differentiations of the R integrals concerned are with respect to ω_{rs} $(r \neq s)$ only, simplification may be introduced by expressing them as differentiations with respect to a, b, c. (3) Absolute moments, when l, m, n are all even, are the same as ordinary moments of the normal distribution. (4) It is easily seen that $(l, 0, 0) = (l)$ and $(l, m, 0) = (l, m)$. (5) We may also note that the moments about zero of the distributions $p(x_1, x_2) = c\,e^{-Q(x_1, x_2)}$ and $p(x_1, x_2, x_3) = ce^{-Q(x_1, x_2, x_3)}$, where Q is a positive definite quadratic form and $x_1, x_2, x_3 > 0$, can be obtained from formulae essentially similar to $[m, n]$ or $[l, m, n]$.

Bivariate results for $[l, m]$:

$$[0, 0] = \frac{1}{2\pi}(\tfrac{1}{2}\pi+\sin^{-1}\rho),$$

$$[1, 0] = \frac{1}{4}\sqrt{\frac{2}{\pi}}(1+\rho),$$

$$[2, 0] = \frac{1}{2\pi}\{\tfrac{1}{2}\pi+\sin^{-1}\rho+\rho\sqrt{(1-\rho^2)}\},$$

$$[1,1] = \frac{1}{2\pi}\{\rho(\tfrac{1}{2}\pi + \sin^{-1}\rho) + \surd(1-\rho^2)\},$$

$$[3,0] = \frac{1}{4}\sqrt{\frac{2}{\pi}}(1+\rho)^2(2-\rho),$$

$$[2,1] = \frac{1}{4}\sqrt{\frac{2}{\pi}}(1+\rho)^2,$$

$$[4,0] = \frac{1}{2\pi}\{3(\tfrac{1}{2}\pi + \sin^{-1}\rho) + \surd(1-\rho^2)(5\rho - 2\rho^3)\},$$

$$[3,1] = \frac{1}{2\pi}\{3\rho(\tfrac{1}{2}\pi + \sin^{-1}\rho) + \surd(1-\rho^2)(2+\rho^2)\},$$

$$[2,2] = \frac{1}{2\pi}\{(1+2\rho^2)(\tfrac{1}{2}\pi + \sin^{-1}\rho) + 3\rho\surd(1-\rho^2)\},$$

$$[5,0] = \frac{1}{4}\sqrt{\frac{2}{\pi}}(1+\rho)^3(8-9\rho+3\rho^2),$$

$$[4,1] = \frac{1}{4}\sqrt{\frac{2}{\pi}}(1+\rho)^3(3-\rho),$$

$$[3,2] = \frac{1}{4}\sqrt{\frac{2}{\pi}}2(1+\rho)^3,$$

$$[6,0] = \frac{1}{2\pi}\{15(\tfrac{1}{2}\pi + \sin^{-1}\rho) + \surd(1-\rho^2)(33\rho - 26\rho^3 + 8\rho^5)\},$$

$$[5,1] = \frac{1}{2\pi}\{15\rho(\tfrac{1}{2}\pi + \sin^{-1}\rho) + \surd(1-\rho^2)(8+9\rho^2-2\rho^4)\},$$

$$[4,2] = \frac{1}{2\pi}\{3(1+4\rho^2)(\tfrac{1}{2}\pi + \sin^{-1}\rho) + \surd(1-\rho^2)(13\rho + 2\rho^3)\},$$

$$[3,3] = \frac{1}{2\pi}\{3\rho(3+2\rho^2)(\tfrac{1}{2}\pi + \sin^{-1}\rho) + \surd(1-\rho^2)(4+11\rho^2)\}.$$

Univariate, bivariate and trivariate absolute moments:

$$(1,0,0) = \sqrt{\frac{2}{\pi}},$$

$$(2,0,0) = 1,$$

$$(1,1,0) = \frac{2}{\pi}\{\surd(1-\rho_{12}^2) + \rho_{12}\sin^{-1}\rho_{12}\},$$

$$(3,0,0) = 2\sqrt{\frac{2}{\pi}},$$

$$(2,1,0) = \sqrt{\frac{2}{\pi}}(1+\rho_{12}^2),$$

$$(1,1,1) = \left(\frac{2}{\pi}\right)^{\frac{3}{2}}\{\omega^{\frac{1}{2}} + \Sigma(\rho_{23} + \rho_{12}\rho_{13})\sin^{-1}\rho_{23\cdot1}\},$$

$$(4,0,0) = 3,$$

$$(3,1,0) = \frac{2}{\pi}\{\surd(1-\rho_{12}^2)(2+\rho_{12}^2) + 3\rho_{12}\sin^{-1}\rho_{12}\},$$

$$(2,2,0) = 1+2\rho_{12}^2,$$

$$(2,1,1) = \frac{2}{\pi}\{(\rho_{23}+2\rho_{12}\rho_{13})\sin^{-1}\rho_{23}+\surd(1-\rho_{23}^2)\,(1+\rho_{12}^2+\rho_{13}^2)\},$$

$$(5,0,0) = 8\sqrt{\frac{2}{\pi}},$$

$$(4,1,0) = \sqrt{\frac{2}{\pi}}\,(3+6\rho_{12}^2-\rho_{12}^4),$$

$$(3,2,0) = 2\sqrt{\frac{2}{\pi}}\,(1+3\rho_{12}^2),$$

$$\begin{aligned}(3,1,1) = \left(\frac{2}{\pi}\right)^{\frac{3}{2}}\{(2+\rho_{12}^2+\rho_{13}^2)\,\omega^{\frac{1}{2}}&+2(\rho_{23}+3\rho_{12}\rho_{13})\sin^{-1}\rho_{23\cdot1}\\ &+[3\rho_{13}(1+\rho_{12}^2)+\rho_{12}\rho_{23}(3-\rho_{12}^2)]\sin^{-1}\rho_{13\cdot2}\\ &+[3\rho_{12}(1+\rho_{13}^2)+\rho_{13}\rho_{23}(3-\rho_{13}^2)]\sin^{-1}\rho_{12\cdot3}\},\end{aligned}$$

$$(2,2,1) = \sqrt{\frac{2}{\pi}}\,(1+2\rho_{12}^2+\rho_{13}^2+\rho_{23}^2-\rho_{13}^2\rho_{23}^2+4\rho_{12}\rho_{13}\rho_{23}),$$

$$(6,0,0) = 15,$$

$$(5,1,0) = \frac{2}{\pi}\{15\rho_{12}\sin^{-1}\rho_{12}+\surd(1-\rho_{12}^2)\,(8+9\rho_{12}^2-2\rho_{12}^4)\},$$

$$(4,2,0) = 3(1+4\rho_{12}^2),$$

$$\begin{aligned}(4,1,1) = \frac{2}{\pi}\Bigg\{3(\rho_{23}+4\rho_{12}\rho_{13})\sin^{-1}\rho_{23}+\surd(1-\rho_{23}^2)\Bigg(&2+8\rho_{12}^2+8\rho_{13}^2+\rho_{23}^2\\ &-4\rho_{12}\rho_{13}\rho_{23}-2\rho_{12}^4-2\rho_{13}^4+4\rho_{12}^2\rho_{13}^2+\frac{\omega^2}{1-\rho_{23}^2}\Bigg)\Bigg\},\end{aligned}$$

$$(3,3,0) = \frac{2}{\pi}\{3\rho_{12}(3+2\rho_{12}^2)\sin^{-1}\rho_{12}+\surd(1-\rho_{12}^2)\,(4+11\rho_{12}^2)\},$$

$$\begin{aligned}(3,2,1) = \frac{2}{\pi}\{&3(\rho_{13}+2\rho_{12}\rho_{23}+2\rho_{12}^2\rho_{13})\sin^{-1}\rho_{13}\\ &+\surd(1-\rho_{23}^2)\,(2+6\rho_{12}^2+\rho_{13}^2+2\rho_{23}^2-2\rho_{13}^2\rho_{23}^2+6\rho_{12}\rho_{13}\rho_{23})\},\end{aligned}$$

$$(2,2,2) = 1+2(\rho_{12}^2+\rho_{13}^2+\rho_{23}^2)+8\rho_{12}\rho_{13}\rho_{23}.$$

Note. The expression ω, included in (1, 1, 1), (3, 1, 1) and (4, 1, 1), is the determinant of the correlation matrix, while the partial correlation coefficients included in (1, 1, 1) and (3, 1, 1) are defined as

$$\rho_{23\cdot1} = (\rho_{23}-\rho_{12}\rho_{13})/\{\surd(1-\rho_{12}^2)\,\surd(1-\rho_{13}^2)\}, \quad \text{etc.}$$

2·7. *Generalization to the multivariate case*

The principle underlying the method given above may now be stated in a general form. The absolute moments of an n-variate normal distribution can be expressed in terms of the derivatives of $n+1$ basic integrals consisting of (i) R_0 and (ii) n integrals of the type R_1, of that order. The latter satisfy a set of n linear equations among themselves and integrals of the type R_0 of the $(n-1)$th order and hence are determinate if the R_0 integral of the $(n-1)$th order is known. For instance, since R_{000} is known, we can find R_{1000}, R_{0100}, R_{0010}, R_{0001}, where

$$R_{1000} = \iiiint_0^\infty x_1\exp\{-\Sigma\omega_{rs}x_rx_s\}\,dx_1\,dx_2\,dx_3\,dx_4 \quad (r,s=1,2,3,4).$$

Thus

$$R_{1000} = \tfrac{1}{2}\omega_{11}^{-1}(\omega_{22}\omega_{33}\omega_{44})^{-\frac{1}{2}}A^{-1}\begin{vmatrix} R_{000}(\alpha_{23},\alpha_{34},\alpha_{24}), & \alpha_{12}, & \alpha_{13}, & \alpha_{14} \\ R_{000}(\alpha_{13},\alpha_{34},\alpha_{14}), & 1, & \alpha_{23}, & \alpha_{24} \\ R_{000}(\alpha_{14},\alpha_{24},\alpha_{12}), & \alpha_{23}, & 1, & \alpha_{34} \\ R_{000}(\alpha_{12},\alpha_{23},\alpha_{13}), & \alpha_{24}, & \alpha_{34}, & 1 \end{vmatrix}, \tag{39}$$

where

$$A = |\alpha_{ij}|, \quad \alpha_{ij} = \omega_{ij}(\omega_{ii}\omega_{jj})^{-\frac{1}{2}}$$

and $R_{000}(a,b,c)$ is the three-dimensional integral

$$\iiint_0^\infty \exp\{-(x^2+y^2+z^2+2axy+2byz+2czx)\}\,dx\,dy\,dz.$$

It seems, however, that the R_0 integral of every order has to be determined directly. Since the evaluation of this integral becomes difficult for $n > 3$, the present approach gives, in finite form, all absolute moments up to the trivariate normal distribution and also the separate contribution of every quadrant, and similar odd-weight results ($l+m+n+k$ odd) for the four-variate case.

3. EXPRESSIONS FOR $[l, m, n, \ldots]$ AND $(l, m, n, \ldots)$ AS POWER-SERIES

3·1. *General theory illustrated in four-variate case*

Since the exact results of §2 are confined to the univariate, bivariate and trivariate cases we are giving here expansions for the absolute moment and the separate contribution of each quadrant in powers of correlation coefficients for the multivariate case. To avoid complicated notation, only the four-variate case is considered, but, as can be seen from the derivation, the method is quite general. Let

$$p(x) = \omega^{-\frac{1}{2}}(2\pi)^{-2}\exp\{-\tfrac{1}{2}\Sigma\omega_{rs}x_r x_s/\omega\} \quad (r, s = 1, 2, 3, 4)$$

be the probability-density function. Then the characteristic function is $\exp\{-\frac{1}{2}\Sigma t_r^2 - \Sigma\rho_{rs}t_r t_s\}$. With a slight change in notation adopted above, the contribution of the positive quadrant to (l_1, l_2, l_3, l_4) is

$$\begin{aligned}
[l_1, l_2, l_3, l_4] &= \iiiint_0^\infty x_1^{l_1}x_2^{l_2}x_3^{l_3}x_4^{l_4}\,p(x)\,dx \\
&= \iiiint_0^\infty x_1^{l_1}x_2^{l_2}x_3^{l_3}x_4^{l_4}\left((2\pi)^{-4}\iiiint_{-\infty}^{\infty}\exp\{-\tfrac{1}{2}\Sigma t_r^2 - \Sigma\rho_{rs}t_r t_s - i\Sigma t_r x_r\}\,dt\right)dx \\
&= (2\pi)^{-4}\iiiint_0^\infty x_1^{l_1}x_2^{l_2}x_3^{l_3}x_4^{l_4}\Bigg[\iiiint_{-\infty}^{\infty}\exp\{-\tfrac{1}{2}\Sigma t_r^2 - i\Sigma t_r x_r\} \\
&\quad\times\left(\Sigma\Sigma\sum_0^\infty\Sigma\Sigma\Sigma\,(-1)^{p+q+r+s+t+u}\frac{\rho_{12}^p\rho_{13}^q\rho_{14}^r\rho_{23}^s\rho_{24}^t\rho_{34}^u}{p!\,q!\,r!\,s!\,t!\,u!}\,t_1^{p+q+r}t_2^{p+s+t}t_3^{q+s+u}t_4^{r+t+u}\right)dt\Bigg]\,dx \\
&= \Sigma\Sigma\sum_0^\infty\Sigma\Sigma\Sigma\,(-1)^{p+q+r+s+t+u}\frac{\rho_{12}^p\rho_{13}^q\rho_{14}^r\rho_{23}^s\rho_{24}^t\rho_{34}^u}{p!\,q!\,r!\,s!\,t!\,u!}\,F_{l_1,\,p+q+r}F_{l_2,\,p+s+t}F_{l_3,\,q+s+u}F_{l_4,\,r+t+u},
\end{aligned} \tag{40}$$

where

$$F_{l,j} = \frac{1}{2\pi}\int_0^\infty x^l\left(\int_{-\infty}^{\infty} t^j\,e^{-\frac{1}{2}t^2 - itx}\,dt\right)dx. \tag{41}$$

An exactly similar procedure gives the absolute moment

$$(l_1,l_2,l_3,l_4) = \iiiint\limits_{-\infty}^{\infty} |\,x_1\,|^{l_1}\,|\,x_2\,|^{l_2}\,|\,x_3\,|^{l_3}\,|\,x_4\,|^{l_4}\,p(x)\,dx$$

$$= \sum\sum\sum_0^{\infty}\sum\sum\sum(-1)^{p+q+r+s+t+u}\frac{\rho_{12}^p\rho_{13}^q\rho_{14}^r\rho_{23}^s\rho_{24}^t\rho_{34}^u}{p!\,q!\,r!\,s!\,t!\,u!}H_{l_1,\,p+q+r}H_{l_2,\,p+s+t}H_{l_3,\,q+s+u}H_{l_4,\,r+t+u}, \tag{42}$$

where

$$H_{l,\,j} = \frac{1}{2\pi}\int_{-\infty}^{\infty}|\,x\,|^l\left(\int_{-\infty}^{\infty}t^j\,e^{-\frac{1}{2}t^2-itx}\,dt\right)dx. \tag{43}$$

To evaluate $F_{l,\,j}$ and $H_{l,\,j}$ we note that

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}t^j\,e^{-\frac{1}{2}t^2-itx}\,dt = \frac{1}{(-i)^j}\frac{1}{\sqrt{(2\pi)}}\frac{d^j}{dx^j}(e^{-\frac{1}{2}x^2}),$$

and therefore

$$F_{l,j} = \frac{1}{(-i)^j\sqrt{(2\pi)}}\int_0^{\infty}x^l\frac{d^j}{dx^j}(e^{-\frac{1}{2}x^2})\,dx. \tag{44}$$

The repeated integration by parts gives

$l \leqslant j$:

$$\left.\begin{aligned} F_{l,\,j} &= (-i)^l\,\tfrac{1}{2}(l!) \quad (j-l=0)\\ &= 0 \quad (j-l \text{ even})\\ &= \frac{(-i)^{l+1}\,l!\,(j-l-1)!}{(2\pi)^{\frac{1}{2}}\,2^{\frac{1}{2}(j-l-1)}\left(\dfrac{j-l-1}{2}\right)!} \quad (j-l \text{ odd}).\end{aligned}\right\} \tag{45}$$

$l \geqslant j$:

$$\begin{aligned} F_{l,\,j} &= \frac{1}{(-i)^j\sqrt{(2\pi)}}\frac{l!}{(l-j)!}\int_0^{\infty}x^{l-j}\,e^{-\frac{1}{2}x^2}\,dx\\ &= \frac{(-i)^j\,(l!)\,2^{\frac{1}{2}(l-j-1)}\,\Gamma\left(\dfrac{l-j+1}{2}\right)}{(2\pi)^{\frac{1}{2}}\,(l-j)!}.\end{aligned} \tag{46}$$

In the case of $H_{l,j}$ it is easily shown that

$$\left.\begin{aligned} H_{l,j} &= 0 \qquad (j \text{ odd})\\ &= 2F_{l,\,j} \quad (j \text{ even}).\end{aligned}\right\} \tag{47}$$

3·2. *Illustrative example*

$$\begin{aligned}[1,1,1,1] = (F_{1,0})^4 &- \frac{(F_{1,1})^2\,(F_{1,0})^2}{1!}\Sigma\rho_{ij} + \frac{(F_{1,2})^2\,(F_{1,0})^2}{2!}\Sigma\rho_{ij}^2\\ &+ \frac{F_{1,2}(F_{1,1})^2\,F_{1,0}}{(1!)^2}\Sigma\rho_{ij}\rho_{ik} + \frac{(F_{1,1})^4}{(1!)^2}\Sigma\rho_{ij}\rho_{kl} - \frac{(F_{1,3})^2\,(F_{1,0})^2}{3!}\Sigma\rho_{ij}^3\\ &- \ldots,\end{aligned} \tag{48}$$

where i, j, k, l are all different. The absolute moment $(1, 1, 1, 1)$ is obtained from (48) by replacing $F_{l,\,j}$ by $H_{l,\,j}$. The substitution for F's and H's gives

$$\begin{aligned}[1,1,1,1] = \frac{1}{4\pi^2} &+ \frac{1}{8\pi}\Sigma\rho_{ij} + \frac{1}{8\pi^2}\Sigma\rho_{ij}^2 - \frac{1}{8\pi}\Sigma\rho_{ij}\rho_{ik} + \tfrac{1}{16}\Sigma\rho_{ij}\rho_{kl}\\ &+ \frac{1}{16\pi}\Sigma\rho_{ij}^2\rho_{kl} + \frac{1}{4\pi^2}\Sigma\rho_{ij}\rho_{ik}\rho_{jk} + \frac{1}{8\pi}\Sigma\rho_{ij}\rho_{ik}\rho_{kl} + \frac{1}{96\pi^2}\Sigma\rho_{ij}^4 + \ldots\end{aligned}$$

and
$$(1,1,1,1) = \frac{4}{\pi^2}\left(1+\tfrac{1}{2}\Sigma\rho_{ij}^2+\Sigma\rho_{ij}\rho_{ik}\rho_{jk}+\tfrac{1}{24}\Sigma\rho_{ij}^4+\tfrac{1}{4}\Sigma\rho_{ij}^2\rho_{ik}^2\right.$$
$$\left.-\tfrac{1}{4}\Sigma\rho_{ij}^2\rho_{kl}^2+\Sigma\rho_{ij}\rho_{ik}\rho_{kl}\rho_{jl}+\ldots\right).$$

It may be observed that $l_1 = l_2 = l_3 = l_4 = 0$ reduces $[l_1, l_2, l_3, l_4]$ to the integral $\iiiint p(x)\,dx$ taken over the positive quadrant (see Moran, 1948; Kendall, 1941, 1945).

4. APPLICATIONS TO THE DISTRIBUTION OF A 'TOTAL ERROR'

4·1. *The moments in special cases*

Let us now consider the distribution of the total error T mentioned in the introduction, confining the discussion to two or three normal and mod-normal component errors. The variances of the component errors will in general be unequal, but the problem can be reduced to the consideration of the following standard forms of T, where, in conformity with the preceding theoretical treatment, the x's have all unit variance:

Case: (i) $x_1+c\,|\,x_2\,|$, (ii) $|\,x_1\,|+c\,|\,x_2\,|$,

(iii) $c_1x_1+c_2\,|\,x_2\,|+c_3\,|\,x_3\,|$, (iv) $c_1\,|\,x_1\,|+c_2\,|\,x_2\,|+c_3\,|\,x_3\,|$.

The distributions of x_1, x_2 and x_1, x_2, x_3 are bivariate and trivariate normal respectively with zero means; c, c_2, c_3 lie between -1 and $+1$; $c_1 = 1$ but is retained as such when dealing with case (iv) below to shorten the writing of moments.

The first four moments of T calculated with the help of the absolute moments are given below for cases (i), (ii) and (iv). In the last case, the moments μ_r' have been given about zero and not the mean, as the expressions for μ_3 and μ_4 are very lengthy:

Case (i). $T = x_1+c\,|\,x_2\,|$.

$$\left.\begin{aligned}
\mu_1' &= c\sqrt{\frac{2}{\pi}},\\
\mu_2 &= 1+c^2\left(1-\frac{2}{\pi}\right),\\
\mu_3 &= \sqrt{\frac{2}{\pi}}\,c\left[3\rho^2+c^2\left(\frac{4}{\pi}-1\right)\right],\\
\mu_4 &= 3[(1+c^2)^2+4c^2\rho^2]-\frac{4}{\pi}[3c^2(1+2\rho^2)+c^4]-\frac{12c^4}{\pi^4}.
\end{aligned}\right\}\tag{49}$$

Case (ii). $T = |\,x_1\,|+c\,|\,x_2\,|$.

$$\left.\begin{aligned}
\mu_1' &= \sqrt{\frac{2}{\pi}}(1+c),\\
\mu_2 &= 1+c^2+\frac{2}{\pi}[2c(\sqrt{(1-\rho^2)}+\rho\sin^{-1}\rho)-(1+c)^2],\\
\mu_3 &= \sqrt{\frac{2}{\pi}}[3c(1+c)\,\rho^2-(1+c^3)]+\left(\frac{2}{\pi}\right)^{\frac{3}{2}}[2(1+c)^3-6c(1+c)\,(\sqrt{(1-\rho^2)}+\rho\sin^{-1}\rho)],\\
\mu_4 &= 3[(1+c^2)^2+4c^2\rho^2]\\
&\quad+\frac{2}{\pi}[6(1-c^2)^2-8(1+c)\,(1+c^3)-12c(1+c)^2\rho^2\\
&\qquad\qquad+4c(1+c^2)\,(\sqrt{(1-\rho^2)}\,(2+\rho^2)+3\rho\sin^{-1}\rho)]\\
&\quad+\left(\frac{2}{\pi}\right)^2[12c(1+c)^2(\sqrt{(1-\rho^2)}+\rho\sin^{-1}\rho)-3(1+c)^4].
\end{aligned}\right\}\tag{50}$$

Case (iv). $T = c_1\,|\,x_1\,| + c_2\,|\,x_2\,| + c_3\,|\,x_3\,|$.

$$\left.\begin{aligned}
\mu_1' &= \sqrt{\frac{2}{\pi}}\,\Sigma c_i,\\
\mu_2' &= \Sigma c_i^2 + \frac{4}{\pi}\,\Sigma c_i c_j(\sqrt{(1-\rho_{ij}^2)} + \rho_{ij}\sin^{-1}\rho_{ij}),\\
\mu_3' &= \sqrt{\frac{2}{\pi}}\,[2\Sigma c_i^3 + 3\Sigma c_i^2 c_j(1+\rho_{ij}^2)]\\
&\quad + \left(\frac{2}{\pi}\right)^{\frac{3}{2}} 6c_1c_2c_3\left[\omega^{\frac{1}{2}} + \Sigma(\rho_{ij}+\rho_{ik}\rho_{jk})\sin^{-1}\left(\frac{\rho_{ij}-\rho_{ik}\rho_{jk}}{\sqrt{(1-\rho_{ik}^2)}\,\sqrt{(1-\rho_{jk}^2)}}\right)\right],\\
\mu_4' &= 3\Sigma c_i^4 + 6\Sigma c_i^2 c_j^2(1+2\rho_{ij}^2)\\
&\quad + \frac{8}{\pi}\{\Sigma c_i^2 c_j[\sqrt{(1-\rho_{ij}^2)}\,(2+\rho_{ij}^2) + 3\rho_{ij}\sin^{-1}\rho_{ij}]\\
&\qquad + 3[\Sigma c_i c_j c_k^2((\rho_{ij}+2\rho_{ik}\rho_{jk})\sin^{-1}\rho_{ij} + \sqrt{(1-\rho_{ij}^2)}\,(1+\rho_{ik}^2+\rho_{jk}^2))]\}\\
&\qquad\qquad (i,j,k = 1,2,3)
\end{aligned}\right\} \quad (51)$$

where ω, used in the expression for μ_3', is the determinant of the correlation matrix of the trivariate normal distribution of x_1, x_2, x_3.

4·2. *Derivation of probability integrals*

To consider the distribution of T, we note that the probability density functions for the first two cases are, respectively,

$$p(x_1, |\,x_2\,|) \equiv p(x,y) = f(x,y;\ \rho) + f(x,y;\ -\rho) \quad (-\infty < x < \infty, 0 < y < \infty) \tag{52}$$

and
$$p(|\,x_1\,|, |\,x_2\,|) \equiv p(x,y) = 2[f(x,y;\ \rho) + 2f(x,y;\ -\rho)] \quad \left(0 < \frac{x}{y} < \infty\right), \tag{53}$$

where
$$f(x,y;\ \rho) = (2\pi)^{-1}(1-\rho^2)^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}(x^2+y^2-2xy\rho)/(1-\rho^2)\}$$

Case (i). $T = x_1 + c\,|\,x_2\,|$.

The probability integral may be written as

$$\begin{aligned}\Pr\{T \leqslant d\} \equiv \Pr\{x+cy \leqslant d\} &= \iint\limits_{x+cy\leqslant d} f(x,y;\ \rho)\,dx\,dy + \iint\limits_{x+cy\leqslant d} f(x,y;\ -\rho)\,dx\,dy\\ &= (I_1 + I_2), \quad \text{where} -\infty < x < \infty, 0 < y < \infty.\end{aligned} \tag{54}$$

By appropriate linear transformations the integrals I_1, I_2 may be expressed in terms of functions of the type

$$p_\rho(h,k) = (2\pi)^{-1}(1-\rho^2)^{-\frac{1}{2}}\int_h^\infty\int_k^\infty \exp\left\{-\frac{1}{2(1-\rho^2)}(x^2+y^2-2\rho xy)\right\}dx\,dy.$$

This function is tabulated in Karl Pearson's (1931) *Tables for Statisticians and Biometricians*, **2**, Tables VIII and IX. Thus, for $d > 0$,

$$I_1 = \tfrac{1}{2} - p_{\rho_1}(d_1, 0), \quad \rho_1 = (\rho+c)/\sqrt{(1+c^2+2c\rho)}, \quad d_1 = d/\sqrt{(1+c^2+2c\rho)}, \tag{55}$$

and for $d < 0 \quad (d = -d', d' > 0)$

$$I_1 = p_{-\rho_1}(d_2, 0), \quad \rho_1 = (\rho+c)/\sqrt{(1+c^2+2c\rho)}, \quad d_2 = d'/\sqrt{(1+c^2+2c\rho)}. \tag{56}$$

I_2 is obtained from (55) and (56) when ρ is replaced by $-\rho$.

Case (ii). $T = |x_1| + c|x_2|$.

$$\Pr\{T \leqslant d\} \equiv \Pr\{x + cy \leqslant d\} = 2\Bigg[\iint\limits_{x+cy\leqslant d} f(x, y;\, \rho)\, dx\, dy + \iint\limits_{x+cy\leqslant d} f(x, y;\, -\rho)\, dx\, dy\Bigg]$$

$$= 2[I_3 + I_4] \quad \left(0 < \begin{matrix} x \\ y \end{matrix} < \infty\right). \tag{57}$$

Here it is necessary to consider $c > 0$, $c < 0$ separately.

$$c > 0\colon I_3 = p_\rho(0, 0) - p_{\rho_1}(d_1, 0) + p_{\rho_2}(d_1, 0), \quad \rho_1 = (\rho + c)/\sqrt{(1 + c^2 + 2c\rho)},$$
$$\rho_2 = -(1 + c\rho)/\sqrt{(1 + c^2 + 2c\rho)}, \quad d_1 = d/\sqrt{(1 + c^2 + 2c\rho)}. \tag{58}$$

$c < 0\; (c = -c', c' > 0)$:

(i) $d > 0\colon I_3 = p_\rho(0, 0) - p_{\rho_3}(d_3, 0), \;\; \rho_3 = (\rho - c')/\sqrt{(1 + c'^2 - 2c'\rho)}, \;\; d_3 = d/\sqrt{(1 + c'^2 - 2c'\rho)},$ (59)

(ii) $d < 0\; (d = -d', d' > 0)$:

$$I_3 = p_{\rho_4}(d_4, 0), \quad \rho_4 = (c'\rho - 1)/\sqrt{(1 + c'^2 - 2c'\rho)}, \quad d_4 = d'/\sqrt{(1 + c'^2 - 2c'\rho)}. \tag{60}$$

The corresponding expressions for I_4 are obtained from (58) to (60) by changing ρ to $-\rho$.

Case (iv). $T = c_1|x_1| + c_2|x_2| + c_3|x_3|$.

It is possible to find the probability integral of T for specific values of c_2, c_3 by numerical quadrature, although the procedure will, of necessity, be long and tedious. Alternatively, a Pearson curve having the exact first four moments may be fitted to give an approximation. To explore this approach the simpler case of uncorrelated x_1, x_2, x_3 was investigated in some detail. In this case the first four moments of T are:

$$\left.\begin{aligned}
\mu_1' &= \sqrt{\frac{2}{\pi}}\,\Sigma c_i,\\
\mu_2 &= \left(1 - \frac{2}{\pi}\right)\Sigma c_i^2\\
\mu_3 &= \sqrt{\frac{2}{\pi}}\left(\frac{4}{\pi} - 1\right)\Sigma c_i^3,\\
\mu_4 &= \left(3 - \frac{4}{\pi} - \frac{12}{\pi^2}\right)\Sigma c_i^4 + 6\left(1 - \frac{2}{\pi}\right)^2 \Sigma c_i^2 c_j^2
\end{aligned}\right\} \quad (i, j = 1, 2, 3). \tag{61}$$

Setting $c_1 = 1$,

$$\left.\begin{aligned}
\beta_1 &= \frac{\dfrac{2}{\pi}\left(1 - \dfrac{4}{\pi}\right)^2}{\left(1 - \dfrac{2}{\pi}\right)^3}\,\frac{(1 + c_2^3 + c_3^3)^2}{(1 + c_2^2 + c_3^2)^3},\\
\beta_2 &= 3 + \frac{\dfrac{8}{\pi}\left(1 - \dfrac{3}{\pi}\right)}{\left(1 - \dfrac{2}{\pi}\right)^2}\,\frac{(1 + c_2^4 + c_3^4)}{(1 + c_2^2 + c_3^2)^2}.
\end{aligned}\right\} \tag{62}$$

Table 1 gives β_1, β_2 for different values of c_2 and c_3. It is seen from the table that the (β_1, β_2) points lie in the rectangle $0 < \beta_1 < 1$, $3 < \beta_2 < 4$ on the β_1, β_2 diagram. When c_2, c_3 are positive the position of the points does not change very much and is always in the Type I region, but if c_2, c_3 are negative it may cross over to Type VI region and then swing on to Type IV region.

4·3. *Adequacy of Pearson-curve approximation tested in special cases*

That a Pearson curve may often give a good approximation to the probability integral (possibly even in the correlated case) is illustrated by the following examples.

Example (i). For the uncorrelated case, with $T = |x_1| + |x_2|$ it is found from (61), (62) by putting $c_2 = 1$, $c_3 = 0$, $\mu_1' = 1{\cdot}5958$, $\sigma(T) = 0{\cdot}853$, $\beta_1 = 0{\cdot}495$, $\beta_2 = 3{\cdot}435$. The upper and lower 5 % points for the corresponding Pearson curve can be obtained from Pearson & Merrington's (1951) table; these values are compared below with those obtained directly by the use of the formula (58), using inverse interpolation:

	Lower 5 %	Upper 5 %
True	0·38	3·17
Pearson fit	0·43	3·17

Table 1. *Values of* β_1 *and* $\beta_2 - 3$ *(bold type) for the distribution of* $T = |x_1| + c_2|x_2| + c_3|x_3|$ *when* x_1, x_2, x_3 *are uncorrelated*

c_3 \ c_2	−1·0	−0·8	−0·6	−0·4	−0·2	0·0	0·2	0·4	0·6	0·8	1·0	c_2 / c_3
−1·0	0·0367 **·290**	0·0141 **·301**	0·0035 **·332**	$0{\cdot}0^{3}40$ **·377**	$0{\cdot}0^{5}75$ **·418**	0·000 **·435**	$0{\cdot}0^{5}75$ **·418**	$0{\cdot}0^{3}40$ **·377**	0·0035 **·332**	0·0141 **·301**	0·0367 **·290**	−1·0
−0·8		$\cdot0^{4}48$ **·304**	·0092 **·334**	·0305 **·385**	·0481 **·435**	·0535 **·455**	·0514 **·435**	·0518 **·385**	·0614 **·334**	·0836 **·304**	·120 **·301**	−0·8
−0·6			·0628 **·370**	·146 **·435**	·217 **·502**	·242 **·531**	·226 **·502**	·203 **·435**	·195 **·370**	·208 **·334**	·240 **·332**	−0·6
−0·4				·328 **·524**	·494 **·620**	·556 **·663**	·511 **·620**	·431 **·524**	·374 **·435**	·356 **·385**	·368 **·377**	−0·4
−0·2					·761 **·748**	·867 **·805**	·786 **·748**	·639 **·620**	·527 **·502**	·473 **·435**	·463 **·418**	−0·2
0·0						0·991 **·869**	0·895 **·805**	0·718 **·663**	0·582 **·531**	0·513 **·455**	0·495 **·435**	0·0
0·2							·811 **·748**	·659 **·620**	·541 **·502**	·483 **·435**	·470 **·418**	0·2
0·4								·548 **·524**	·462 **·435**	·422 **·385**	·419 **·377**	0·4
0·6									·399 **·370**	·370 **·334**	·370 **·332**	0·6
0·8										·342 **·304**	·340 **·301**	0·8
1·0											·330 **·290**	1·0

Example (2). The mean successive difference with the sample size three, i.e.

$$T = \tfrac{1}{2}\{|x_1 - x_2| + |x_2 - x_3|\},$$

is an illustration of a correlated total error. In this case it can be shown using (50) that

$$\mu_1' = 1{\cdot}1284, \quad \sigma(T) = 0{\cdot}670, \quad \beta_1 = 1{\cdot}036, \quad \beta_2 = 4{\cdot}296.$$

The lower and upper 5 % points for the corresponding Pearson curve as obtained from the Pearson-Merrington table are compared below with those obtained for the exact distribution by using (58):

	Lower 5 %	Upper 5 %
True	0·27	2·40
Pearson fit	0·28	2·40

Example (3). The distribution of the mean deviation has been obtained by Godwin (1945). Considering the case of the sample size three, we have, using (51),

$$\mu_1' = 0{\cdot}6515, \quad \sigma = 0{\cdot}342, \quad \beta_1 = 0{\cdot}417, \quad \beta_2 = 3{\cdot}286.$$

The following is a comparison of the actual percentage points with those obtained from the Pearson-Merrington table:

	Lower 0·5 %	Lower 5 %	Upper 5 %	Upper 0·5 %
True	0·052	0·166	1·276	1·703
Pearson fit	−0·009	0·157	1·270	1·696

It is to be noted that the last figure in the second line is not reliable since the Pearson-Merrington table gives only two places of decimals for the standardized deviate.

Before concluding I should like to express my grateful thanks to Dr H. O. Hartley for his advice during these investigations and to Prof. E. S. Pearson for helpful suggestions in the writing of this paper.

REFERENCES

ARLEY, N. & HALD, A. (1950). *Math. Tidsskr.* B, **86**.
GODWIN, H. J. (1945). *Biometrika*, **33**, 265.
KENDALL, M. G. (1941). *Biometrika*, **32**, 196.
KENDALL, M. G. (1945). *J. R. Statist. Soc.* **108**, 93.
MORAN, P. A. P. (1948). *Biometrika*, **35**, 203.
NABEYA, S. (1951). *Ann. Inst. Statist. Math.* **3**, 2.
NABEYA, S. (1952). *Ann. Inst. Statist. Math.* **4**, 15.
PEARSON, E. S. & MERRINGTON, M. (1951). *Biometrika*, **38**, 4.
PEARSON, K. (1931). *Tables for Statisticians and Biometricians*, **2**. London: Biometrika Trust.
TRICOMI, F. (1936). *G. Ist. Ital. Attuari*, **7**, 280.

ON THE RANGE OF PARTIAL SUMS OF A FINITE NUMBER OF INDEPENDENT NORMAL VARIATES

By A. A. ANIS and E. H. LLOYD

Imperial College, London

The properties of the range of the partial sums $S_1, S_2, \ldots, S_n$ of the independent variates $X_1, X_2, \ldots, X_n$ are important in the theory of storage capacity. The asymptotic distribution, for large n, is known (Feller, 1951). The present paper discusses the mean value of the range for finite n, in the case where the X_i are standard normal variates. This is shown to depend on the integral

$$(2\pi)^{-\frac{1}{2}r}\int_0^\infty \ldots (r) \ldots \int_0^\infty \exp -\tfrac{1}{2}\{y_1^2+(y_1-y_2)^2+\ldots+(y_{r-1}-y_r)^2+y_r^2\}\, dy_1 \ldots dy_r,$$

whose value is found to be $(r+1)^{-\frac{3}{2}}$.

The expected value of the range is shown to be

$$\sqrt{\left(\frac{2}{\pi}\right)} \sum_1^{n-1} r^{-\frac{1}{2}}.$$

1. Introduction

The present investigation was suggested by the problem of planning the storage capacity of a reservoir, where one would like to know the distribution of the water level over a given number n of years. The level after r years may be regarded as the sum of r annual increments, and we may approximate this real problem by the ideal one where the annual increments are independent variates with a common distribution. Applications to other storage problems are obvious.

In the case of very large n the asymptotic range of the consequent random walk has been discussed by Feller (1951). In the present paper we are concerned with an exact formulation for the case of finite n, taking the increments to be standard normal variates, and for this case we obtain a simple expression for the mean range.

The paper falls into two parts. In the first (§§ 2–5), the problem is reduced to that of evaluating the integral

$$\int_0^\infty \ldots (r) \ldots \int_0^\infty \exp\left\{\sum_1^{r-1} y_i y_{i+1} - \sum_1^r y_i^2\right\} dy_1 \ldots dy_r.$$

In the second part (§ 6) this integral is shown to have the value $(\sqrt{(2\pi)})^r (r+1)^{-\frac{3}{2}}$, whence the expected value of the range, over n years, is found to be

$$\sqrt{\left(\frac{2}{\pi}\right)} \sum_1^{n-1} r^{-\frac{1}{2}}.$$

2. Statement of the problem

We consider n independent variates $X_1, X_2, \ldots, X_n$, each with the same probability density function $\phi(x)$, and their partial sums

$$S_r = X_1 + X_2 + \ldots + X_r \quad (r = 1, 2, \ldots, n).$$

We let

$$U_n = \operatorname*{Max}_r \{S_r\} \quad \text{and} \quad L_n = \operatorname*{Min}_r \{S_r\};$$

then the mean range is $\mathscr{E}(U_n - L_n) = \mathscr{E}(U_n) - \mathscr{E}(L_n)$. Let $f_n(x)$ be the probability density function of U_n, and let $F_n(x)$ and $1 - G_n(x)$ respectively be the distribution functions of U_n and L_n, so that

$$F_n(x) = \Pr(U_n \leqslant x) \quad \text{and} \quad G_n(x) = \Pr(L_n \geqslant x).$$

We have

$$F_n(y) = \underset{\substack{\sum_1^r x_i \leqslant y, \\ r=1,2,\ldots,n}}{\int \ldots (n) \ldots \int} \prod_{i=1}^{n} \phi(x_i)\, dx_i$$

and

$$G_n(y) = \underset{\substack{\sum_1^r x_i \geqslant y, \\ r=1,2,\ldots,n}}{\int \ldots (n) \ldots \int} \prod_{i=1}^{n} \phi(x_i)\, dx_i.$$

This latter may be alternatively written in the form

$$G_n(-y) = \underset{\substack{\sum_1^n x_i \leqslant y, \\ r=1,2,\ldots,n}}{\int \ldots (n) \ldots \int} \prod_{i=1}^{n} \phi(-x_i)\, dx_i,$$

corresponding to the observation that $\underset{r}{\text{Min}}\{S_r\} = -\underset{r}{\text{Max}}\{-S_r\}$. The functions F_n and G_n are thus integrals of the same type with identical boundary conditions. If $\phi(x)$ is an even function it follows of course that $F_n(y) = G_n(-y)$.

3. $\mathscr{E}(U_n)$ AS A LINEAR FORM IN THE $F_i(0)$

If we make the transformation

$$\sum_1^r x_i = y_0 - y_r \quad \text{or} \quad x_r = y_{r-1} - y_r,$$

the above expression for F_n becomes

$$F_n(y_0) = \int_0^\infty \ldots (n) \ldots \int_0^\infty \prod_{r=1}^{n} \phi(y_{r-1} - y_r)\, dy_r.$$

It will be convenient to define $F_0(x) \equiv 1$. It is clear from this formulation that

$$F_{n+1}(y_0) = \int_0^\infty \phi(y_0 - y_1)\, F_n(y_1)\, dy_1, \tag{3·1}$$

whence, on differentiating and then integrating by parts,

$$f_{n+1}(y_0) = F_n(0)\,\phi(y_0) + \int_0^\infty \phi(y_0 - y_1) f_n(y_1)\, dy_1. \tag{3·2}$$

Using this as a reduction formula we are led to

$$f_{n+1}(y_0) = \sum_{r=0}^{n} F_{n-r}(0)\, h_r(y_0), \tag{3·3}$$

where

$$h_r(y_0) = \int_0^\infty (r) \int_0^\infty \phi(y_0 - y_1)\,\phi(y_1 - y_2) \ldots \phi(y_{r-1} - y_r)\,\phi(y_r)\, dy_1 \ldots dy_r \quad (r = 1, 2, \ldots, n),$$

and $h_0(y) = \phi(y)$.

The expected value μ_{n+1} of U_{n+1} is equal to $\int_{-\infty}^{+\infty} x f_{n+1}(x)\,dx$, and, using (3·3), we obtain

$$\mu_{n+1} = \sum_{r=0}^{n} q_r F_{n-r}(0), \tag{3·4}$$

where

$$q_r = \int_{-\infty}^{+\infty} y_0 h_r(y_0)\,dy_0.$$

Since

$$\int_{-\infty}^{+\infty} y_0 \phi(y_0 - y_1)\,dy_0 = y_1$$

it is easily seen that

$$q_r = \int_0^\infty \dots (r) \dots \int_0^\infty y_1 \phi(y_1 - y_2) \dots \phi(y_{r-1} - y_r)\, \phi(y_r)\, dy_1 \dots dy_r \quad (r = 1, 2, \dots n), \tag{3·5}$$

with $q_0 = 0$.

4. AN IDENTITY IN THE DISTRIBUTION FUNCTIONS

If we consider a particular S_r we have

$$S_{r-i} = S_r - (X_r + X_{r-1} + \dots + X_{r-i+1}) = S_r - T_i, \text{ say}, \quad (i = 1, 2, \dots, r-1),$$

and

$$S_{r+j} = S_r + (X_{r+1} + X_{r+2} + \dots + X_{r+j}) = S_r + R_j, \text{ say}, \quad (j = 1, 2, \dots, n-r).$$

If S_r is the greatest of the partial sums we must have

$$T_i \geqslant 0 \quad (i = 1, 2, \dots, r-1) \quad \text{and} \quad R_j \leqslant 0 \quad (j = 1, 2, \dots, n-r),$$

and, in particular, $\operatorname*{Min}_i \{T_i\} \geqslant 0$, $\operatorname*{Max}_j \{R_j\} \leqslant 0$. Now T_i and R_j are independent partial sums so that

$$\Pr(S_r = \operatorname*{Max}_k \{S_k\}) = G_{r-1}(0)\, F_{n-r}(0).$$

Further, since some S_r must be the largest, the probability being 0 that any two partial sums are equal, we have

$$\sum_{r=1}^{n} G_{r-1}(0)\, F_{n-r}(0) = 1.$$

Where $\phi(x)$ is an even function this becomes

$$\sum_{r=1}^{n} F_{r-1}(0)\, F_{n-r}(0) = 1. \tag{4·1}$$

5. THE SPECIAL CASE OF NORMAL X_i

We now turn to the special case where the X_i are standard normal variates, so that $\phi(u) = \phi(-u)$ and $\phi'(u) = -u\phi(u)$. Applying this latter result after differentiating (3·1) we get, with a slight change of notation,

$$f_{n+1}(y) = \int_0^\infty (x - y)\, \phi(y - x)\, F_n(x)\, dx.$$

If we multiply (3·1) by y and add to this the last result, we find

$$f_{n+1}(y) + y F_{n+1}(y) = \int_0^\infty x \phi(y - x)\, F_n(x)\, dx. \tag{5·1}$$

Taking $n = 0, 1, \ldots, r-1$ in this formula, and applying the results to the reduction of the expression (3·5) for q_r, we find

$$q_r = \int_0^\infty (r-1) \int_0^\infty f_1(y_2)\, \phi(y_2-y_3) \ldots \phi(y_{r-1}-y_r)\, \phi(y_r)\, dy_2 \ldots dy_r$$

$$+ \ldots + \int_0^\infty \int_0^\infty f_{r-2}(y_{r-1})\, \phi(y_{r-1}-y_r)\, \phi(y_r)\, dy_{r-1}\, dy_r$$

$$+ \int_0^\infty f_{r-1}(y_r)\, \phi(y_r)\, dy_r + \int_0^\infty y_r\, F_{r-1}(y_r)\, \phi(y_r)\, dy_r. \tag{5·2}$$

The last term in this expression equals $-\int_0^\infty F_{r-1}(y_r)\, \phi'(y_r)\, dy_r$, and this may be integrated to give

$$F_{r-1}(0)\, \phi(0) + \int_0^\infty f_{r-1}(y_r)\, \phi(y_r)\, dy_r.$$

This last integral is identical with the penultimate term of (5·2). If in this we substitute for $f_{r-1}(y_r)$ according to (3·2), the integral becomes

$$\int_0^\infty f_{r-1}(y_r)\, \phi(y_r)\, dy_r = F_{r-2}(0) \int_0^\infty \phi^2(y_r)\, dy_r + \int_0^\infty \int_0^\infty f_{r-2}(y_{r-1})\, \phi(y_{r-1}-y_r)\, \phi(y_r)\, dy_{r-1}\, dy_r.$$

The last integral here again coincides with a term of (5·2). Continuing in this way we finally obtain for (5·2)

$$q_r = \sum_{s=1}^{r} s b_s F_{r-s}(0) \quad (r = 1, 2, \ldots, n), \tag{5·3}$$

where

$$b_s = \int_0^\infty (s-1) \int_0^\infty \phi(y_{r-s+2})\, \phi(y_{r-s+2}-y_{r-s+3}) \ldots \phi(y_{r-1}-y_r)\, \phi(y_r)\, dy_{r-s+2} \ldots dy_r$$

$$(s = 3, 4, \ldots, n), \tag{5·4}$$

with

$$b_1 = \phi(0) \quad \text{and} \quad b_2 = \int_0^\infty \phi^2(y)\, dy.$$

Using the expansion (3·4) for μ_{n+1} we obtain

$$\mu_{n+1} = \sum_{r=1}^{n} \sum_{s=1}^{r} s b_s F_{r-s}(0)\, F_{n-r}(0),$$

which, with the aid of the identity (4·1), reduces to

$$\mu_{n+1} = \sum_{s=1}^{n} s b_s. \tag{5·5}$$

The problem is thus reduced to that of evaluating the b_s. It is convenient to rewrite (5·4) in the form

$$b_s = (2\pi)^{-\frac{1}{2}s}\, c_{s-1} \quad (s = 1, 2, \ldots, n), \tag{5·6}$$

where, explicitly now,

$$c_0 = 1, \quad c_1 = \int_0^\infty e^{-y^2}\, dy,$$

$$c_s = \int_0^\infty (s) \int_0^\infty \exp - \tfrac{1}{2}\{y_1^2 + (y_1-y_2)^2 + \ldots + (y_{s-1}-y_s)^2 + y_s^2\}\, dy_1 \ldots dy_s \quad (s = 2, 3, \ldots, n-1). \tag{5·7}$$

6. Evaluation of the c_s

The remaining part of the paper is concerned with evaluating the c_s. We use matrix methods, with the convention that lower- and upper-case bold-face letters denote column vectors and square matrices respectively. An expression such as

$$\boldsymbol{\lambda} \geqslant \mathbf{0},$$

where $\boldsymbol{\lambda}$ is a vector of order k, means

$$\lambda_i \geqslant 0 \quad (i = 1, 2, \ldots, k).$$

We write (5·7) as

$$c_r = \int_{\mathbf{y} \geqslant \mathbf{0}}^{(r)} \int e^{-\frac{1}{2}\mathbf{y}'\mathbf{A}\mathbf{y}}\, dy_1 \ldots dy_r, \tag{6·1}$$

where $\mathbf{A}$ is the $(r \times r)$ matrix

$$\mathbf{A} = \begin{pmatrix} 2 & -1 & 0 & \ldots & \ldots & 0 \\ -1 & 2 & -1 & 0 & \ldots & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ \multicolumn{6}{c}{\dotfill} \\ 0 & \ldots & 0 & -1 & 2 & -1 \\ 0 & \ldots & \ldots & 0 & -1 & 2 \end{pmatrix}.$$

This matrix is symmetric and positive-definite, and may therefore be resolved into triangular factors:

$$\mathbf{A} = \mathbf{T}'\mathbf{T},$$

where $\mathbf{T}$ is upper triangular. It is easily found that

$$\mathbf{T} = \begin{pmatrix} \sqrt{\frac{2}{1}} & -\sqrt{\frac{1}{2}} & 0 & 0 & \ldots & 0 \\ 0 & \sqrt{\frac{3}{2}} & -\sqrt{\frac{2}{3}} & 0 & \ldots & 0 \\ 0 & 0 & \sqrt{\frac{4}{3}} & -\sqrt{\frac{3}{4}} & \ldots & 0 \\ \multicolumn{6}{c}{\dotfill} \\ 0 & \ldots & \ldots & 0 & 0 & -\sqrt{\frac{r+1}{r}} \end{pmatrix}.$$

The exponent in the integrand of c_r in (6·1) thus becomes

$$\mathbf{y}'\mathbf{A}\mathbf{y} = \mathbf{y}'\mathbf{T}'\mathbf{T}\mathbf{y} = \mathbf{u}'\mathbf{u},$$

say, where

$$\mathbf{u} = \mathbf{T}\mathbf{y}.$$

Let us make this substitution in the integral. The Jacobian is

$$1/|\,\mathbf{A}\,| = 1/\sqrt{(r+1)}.$$

Then (6·1) becomes

$$(r+1)^{\frac{1}{2}}\, c_r = \int_{\mathbf{T}^{-1}\mathbf{u} \geqslant \mathbf{0}}^{(r)} \int e^{-\frac{1}{2}\mathbf{u}'\mathbf{u}}\, du_1 \ldots du_r. \tag{6·2}$$

The region of integration may be written in the form

$$\mathbf{u} = \mathbf{T}\mathbf{z} \quad (\mathbf{z} \geqslant \mathbf{0}),$$

or, slightly more conveniently for our purpose, as

$$\mathbf{u} = \mathbf{L}\mathbf{x} \quad (\mathbf{x} \geqslant \mathbf{0}), \tag{6·3}$$

where

$$\begin{aligned} \mathbf{L} &= \mathbf{T}/\sqrt{2} \\ &= (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \ldots, \boldsymbol{\lambda}_r), \end{aligned}$$

and the columns $\boldsymbol{\lambda}_i$ of $\mathbf{L}$ are defined explicitly by

$$\boldsymbol{\lambda}_1' = (1, \mathbf{0}_{r-1}),$$

$$\boldsymbol{\lambda}_s' = \left(\mathbf{0}_{s-2}, -\sqrt{\frac{s-1}{2s}}, \sqrt{\frac{s+1}{2s}}, \mathbf{0}_{r-s}\right) \quad (s = 2, 3, \ldots, r).$$

Here $\mathbf{0}_i$ denotes the zero vector of order i. We note that the $\boldsymbol{\lambda}_i$ are normalized, i.e.

$$\boldsymbol{\lambda}_i'\boldsymbol{\lambda}_i = 1 \quad (i = 1, 2, \ldots, r).$$

We now introduce a vector $\boldsymbol{\lambda}_{r+1}$ defined by the relation

$$\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2 + \ldots + \boldsymbol{\lambda}_{r+1} = \mathbf{0}.$$

It is easily verified that we then have

$$\left.\begin{array}{ll} \boldsymbol{\lambda}_i'\boldsymbol{\lambda}_{i+1} = -\tfrac{1}{2} & \\ \boldsymbol{\lambda}_i'\boldsymbol{\lambda}_{i+j} = 0 & (j > 1) \end{array}\right\} \quad (i = 1, 2, \ldots, r+1),$$

the suffices being reduced modulo $r+1$ when appropriate, and that the new vector $\boldsymbol{\lambda}_{r+1}$ is normalized.

We now introduce the following $(r \times r)$ matrices:

$$\mathbf{L}_1 = (\boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3, \ldots, \boldsymbol{\lambda}_{r+1})$$
$$\mathbf{L}_2 = (\boldsymbol{\lambda}_3, \boldsymbol{\lambda}_4, \ldots, \boldsymbol{\lambda}_{r+1}, \boldsymbol{\lambda}_1)$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$\mathbf{L}_{r+1} = (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \ldots, \boldsymbol{\lambda}_r),$$

the last of these being the matrix $\mathbf{L}$ of (6·3). It may be seen that

$$\mathbf{L}_1'\mathbf{L}_1 = \mathbf{L}_2'\mathbf{L}_2 = \ldots = \mathbf{L}_{r+1}'\mathbf{L}_{r+1} (= \tfrac{1}{2}\mathbf{A}),$$

whence the $\mathbf{L}_i$ may be shown to be orthogonal transforms of one another; for let

$$\mathbf{P}\mathbf{L}_i = \mathbf{L}_j \quad (i \neq j);$$

then

$$\mathbf{L}_i'\mathbf{P}'\mathbf{P}\mathbf{L}_i = \mathbf{L}_j'\mathbf{L}_j$$
$$= \mathbf{L}_i'\mathbf{L}_i,$$

so that

$$\mathbf{P}'\mathbf{P} = \mathbf{L}_i'^{-1}\mathbf{L}_i'\mathbf{L}_i\mathbf{L}_i^{-1} = \mathbf{I},$$

where $\mathbf{I}$ denotes the unit matrix. Hence $\mathbf{P}$ is orthogonal.

Since the integrand of c_r is invariant under orthogonal transformations of $\mathbf{u}$, we may therefore describe the region of integration (6·3) in any of the alternative forms

$$\mathscr{R}_i: \quad \mathbf{u} = \mathbf{L}_i\mathbf{y} \quad (\mathbf{y} \geqslant \mathbf{0};\ i = 1, 2, \ldots, r+1).$$

Hence

$$(r+1)^{\frac{1}{2}} c_r = \int\underset{\mathscr{R}_1}{(r)}\int = \int\underset{\mathscr{R}_2}{(r)}\int = \ldots = \int\underset{\mathscr{R}_{r+1}}{(r)}\int = \frac{1}{r+1}\sum_{i=1}^{r+1}\int\underset{\mathscr{R}_i}{(r)}\int, \qquad (6{\cdot}4)$$

where in each case the integrand is that of (6·2).

We now prove

(i) that the $r+1$ regions $\mathscr{R}_i$ are non-overlapping, so that the sum of integrals in (6·4) may be replaced by a single r-fold integral over the union of the $\mathscr{R}_i$; and

(ii) that the regions $\mathscr{R}_i$ are exhaustive, so that their union is the whole space.

Integration is then immediate.

(i) *The regions are non-overlapping*

Consider, for example, the regions $\mathscr{R}_1$ and $\mathscr{R}_2$. If $\mathbf{u}$ is any point common to them, we shall have

$$\mathbf{u} = \mathbf{L}_1\mathbf{x} = \mathbf{L}_2\mathbf{y} \quad (\mathbf{x} \geqslant \mathbf{0}, \mathbf{y} \geqslant \mathbf{0}).$$

Then

$$\mathbf{L}_1'\mathbf{L}_1\mathbf{x} = \mathbf{L}_1'\mathbf{L}_2\mathbf{y},$$

or, explicitly,

$$\begin{pmatrix} 1 & -\frac{1}{2} & 0 & \dots & \dots & \dots & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & \dots & \dots & 0 \\ 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \dots & 0 & -\frac{1}{2} & 1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} -\frac{1}{2} & 0 & \dots & \dots & \dots & -\frac{1}{2} \\ 1 & -\frac{1}{2} & 0 & \dots & \dots & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & -\frac{1}{2} & 1 & -\frac{1}{2} \end{pmatrix} \mathbf{y}.$$

The solution of this set of equations is

$$x_s = y_{s-1} - y_r \quad (s = 2, 3, \dots, r),$$
$$x_1 = -y_r.$$

Since by hypothesis the elements of $\mathbf{x}$ and $\mathbf{y}$ are all non-negative, it follows that

$$x_1 = y_r = 0.$$

Any common points $\mathbf{u}$ thus lie in an $(r-1)$-dimensional subspace, whose measure in r-space is of course zero. Speaking geometrically, we may say that the regions have common boundaries but are non-intersecting.

A precisely similar argument applies to all other pairs of regions $\mathscr{R}_i, \mathscr{R}_j$. Hence

$$c_r = (r+1)^{-\frac{3}{2}} \int \underset{\mathscr{R}}{(r)} \int e^{-\frac{1}{2}\mathbf{u}'\mathbf{u}}\, du_1 \dots du_r, \tag{6·5}$$

where the region of integration $\mathscr{R}$ is the union of the regions $\mathscr{R}_i$:

$$\mathscr{R} = \bigcup_{i=1}^{r+1} \mathscr{R}_i.$$

(ii) *The $r+1$ regions $\mathscr{R}_i$ are exhaustive*

We have to prove that every r-dimensional vector is of the form

$$\mathbf{u} = \mathbf{L}_r\mathbf{y} \quad (\mathbf{y} \geqslant \mathbf{0}),$$

for some r. Now every set of r of the basis vectors $\boldsymbol{\lambda}_i$ is linearly independent, so that we may always write

$$\mathbf{u} = \mathbf{L}_1\mathbf{x} = \mathbf{L}_2\mathbf{y} = \dots = \mathbf{L}_{r+1}\mathbf{w},$$

for any $\mathbf{u}$, with appropriate values of $\mathbf{x}, \mathbf{y}, \dots, \mathbf{w}$. Let us write these representations in the forms

$$\mathbf{u} = \boldsymbol{\lambda}_2 x_2 + \dots + \boldsymbol{\lambda}_{r+1} x_{r+1},$$
$$\dots\dots\dots\dots\dots\dots\dots\dots$$
$$\mathbf{u} = \boldsymbol{\lambda}_1 w_1 + \dots + \boldsymbol{\lambda}_r w_r.$$

Our proof is by induction. Assume for a particular value of s that, if in one of these representations the number of negative coefficients does not exceed s, then $\mathbf{u}$ lies in one of the regions $\mathscr{R}_i$. Now consider the case where in some representation (say the last) there are $s+1$ negative coefficients. Suppose, as we may with no loss of generality, that these are $w_1, \ldots, w_{s+1}$, so that

$$w_i = -\omega_i \quad (i = 1, 2, \ldots, s+1),$$

and

$$w_j = +\omega_j \quad (j = s+2, \ldots, r),$$

where *all* the ω_k are non-negative.

We replace $\boldsymbol{\lambda}_1$ by $-\boldsymbol{\lambda}_2 - \ldots - \boldsymbol{\lambda}_{r+1}$. The representation then becomes

$$\begin{aligned}\mathbf{u} &= -\omega_1\boldsymbol{\lambda}_1 - \ldots - \omega_{s+1}\boldsymbol{\lambda}_{s+1} + w_{s+2}\boldsymbol{\lambda}_{s+2} + \ldots + w_r\boldsymbol{\lambda}_r \\ &= (\omega_1 - \omega_2)\boldsymbol{\lambda}_2 + \ldots + (\omega_1 - \omega_{s+1})\boldsymbol{\lambda}_{s+1} + (\omega_1 + \omega_{s+2})\boldsymbol{\lambda}_{s+2} + \ldots + (\omega_1 + \omega_r)\boldsymbol{\lambda}_r + \omega_1\boldsymbol{\lambda}_{r+1}.\end{aligned}$$

This is a representation of $\mathbf{u}$ in which the only coefficients that can possibly be negative are those attached to $\boldsymbol{\lambda}_2, \ldots, \boldsymbol{\lambda}_{s+1}$. Thus there are at most s negative coefficients, whence, if our initial assumption were valid, it would follow that the point under discussion must belong to one of the regions $\mathscr{R}_i$.

Now the assumption is certainly valid for $s = 0$, and is therefore true in general for $s = 0, 1, \ldots, r$. Thus any r-dimensional vector whose negative coefficients in some representation do not exceed r in number, must belong to one of the regions $\mathscr{R}_i$, and since this clearly exhausts all the possible sign variations of the r coefficients our lemma is proved.

It follows then that the union of the $\mathscr{R}_i$ is the whole space. Hence (6·5) becomes

$$\begin{aligned}(r+1)^{\frac{1}{2}}\, c_r &= \int_{-\infty}^{+\infty} (r) \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\mathbf{u}'\mathbf{u}}\, du_1 \ldots du_r \\ &= \left\{\int_{-\infty}^{+\infty} e^{-\frac{1}{2}u^2}\, du\right\}^r = (\sqrt{(2\pi)})^r.\end{aligned}$$

Referring back to (5·6) we have

$$b_r = (2\pi)^{-\frac{1}{2}r}\, c_{r-1} = (2\pi r^3)^{-\frac{1}{2}},$$

and finally, by (5·5),

$$\sqrt{(2\pi)}\,\mu_{n+1} = \sum_{r=1}^{n} r^{-\frac{1}{2}}.$$

In conclusion, we note that the asymptotic value of the mean range, for large n, is $2\sqrt{\left(\frac{2}{\pi}\right)}\, n^{\frac{1}{2}}$, in agreement with Feller's results.

REFERENCE

FELLER, W. (1951). The asymptotic distribution of the range of sums of independent random variables. *Ann. Math. Statist.* **22**, 427.

NOTE ON 'THE JACOBIANS OF CERTAIN MATRIX TRANSFORMATIONS USEFUL IN MULTIVARIATE ANALYSIS'

BY INGRAM OLKIN

Michigan State College[*]

SUMMARY AND INTRODUCTION

General techniques for the evaluation of the Jacobian of a matrix transformation were outlined by Deemer & Olkin (1951). In particular, (*a*) the Jacobian of a non-linear transformation is equal to the Jacobian of the linear transformation in the differentials, (*b*) by the introduction of suitable variables, the Jacobian is equal to a product of Jacobians which are easily calculated. This note is concerned with a number of transformations not considered in the earlier paper. Certain of the transformations considered will be recognized as extensions of those given previously, while others are quite different. The notation and results of the first paper are assumed. All matrices in this note are square of order p.

THE JACOBIANS OF MATRIX TRANSFORMATIONS

THEOREM 1. The J of the transformation

$$\tilde{Y} = \tilde{X}\tilde{A} \quad \text{is} \quad D(\tilde{X};\, \tilde{Y}) = \prod_1^p a_{ii}^{p-i+1}.$$

Proof. $y_{ij} = \sum_{k=j}^{i} X_{ik} a_{kj} \,(i \geqslant j)$. The scheme of coefficients is triangular with diagonal elements $a_{kk}\,(p-k+1)$ times from

$$\partial y_{ik}/\partial x_{ik} \quad (i = k, \ldots, p;\; k = 1, 2, \ldots, p).$$

Hence the determinant is $\prod_1^p a_{ii}^{p-i+1}$.

THEOREM 2. The J of the transformation

$$\bar{Y} = \tilde{X}'\tilde{A} + \tilde{A}'\tilde{X} \quad \text{is} \quad D(\tilde{X};\, \bar{Y}) = 2^p \prod_1^p a_{ii}^{i}.$$

Proof. Pre- and post-multiply by $\tilde{A}'^{-1}$ and $\tilde{A}^{-1}$, respectively, and let $\bar{Z} = \tilde{A}'^{-1}\bar{Y}\tilde{A}^{-1}$, $\tilde{U} = \tilde{X}\tilde{A}^{-1}$. Then $\bar{Z} = \tilde{U} + \tilde{U}'$ and

$$D(\tilde{X};\, \bar{Y}) = D(\tilde{X};\, \tilde{U})\, D(\tilde{U};\, \bar{Z})\, D(\bar{Z};\, \bar{Y}).$$

We can now evaluate each component

$$D(\tilde{X};\, \tilde{U}) = \prod_1^p a_{ii}^{-(p-i+1)}, \quad D(\bar{Z};\, \bar{Y}) = |\tilde{A}\,|^{p+1} = \prod_1^p a_{ii}^{p+1}, \quad D(\tilde{U};\, \bar{Z}) = 2^p.$$

COROLLARY 2. The J of the transformation

$$\bar{V} = \tilde{T}'\tilde{T} \quad \text{is} \quad D(\tilde{T};\, \bar{V}) = 2^p \prod_1^p t_{ii}^{i}.$$

Proof. Taking differentials of $\bar{V} = \tilde{T}'\tilde{T}$ and making the correspondence $(d\bar{V}) = \bar{Y}$, $(d\tilde{T}) = \tilde{X}$, $\tilde{T} = \tilde{A}$, of Theorem 2, the corollary follows.

---

[*] The work reported here was completed at the University of North Carolina and was supported in part by the Office of Naval Research.

THEOREM 3. The J of the transformation $\bar{V} = D_x \bar{R} D_x$, where $r_{ii} = 1$, is

$$D(x, \bar{R};\, \bar{V}) = 2^p \prod_1^p x_i^p.$$

Proof. $v_{ii} = x_i^2$, $v_{ij} = x_i x_j r_{ij}\,(i \neq j)$. Hence the scheme of coefficients is

$$\begin{array}{c|cc} & x_i & r_{ij} \\ \hline v_{ii} & 2D_1 & 0 \\ v_{ij} & L & D_2 \end{array}$$

where $D_1 : p \times p$ and $D_2 : p(p-1)/2 \times p(p-1)/2$ are diagonal matrices with elements $2x_1, \ldots, 2x_p$ and $x_1x_2, x_1x_3, \ldots, x_{p-1}x_p$, respectively. The matrix L does not affect the value of the determinant, which is equal to

$$2^p \prod_1^p x_i \prod_{i<j} x_i x_j = 2^p \prod_1^p x_i^p.$$

THEOREM 4. The J of the transformation $\tilde{T} = D_x \tilde{U}$, where $\sum_{j=1}^{i} u_{ij}^2 = 1\,(i = 1, \ldots, p)$ is

$$D(x, \tilde{U};\, \tilde{T}) = \prod_1^p x_i^{i-1} u_{ii}^{-1}.$$

Proof.
$$t_{ii} = x_i u_{ii} = x_i \left(1 - \sum_{j=1}^{i-1} u_{ij}^2\right)^{\frac{1}{2}}, \quad t_{ij} = x_i u_{ij}\,(i > j).$$

Hence the scheme of coefficients is

$$\begin{array}{c|cc} & x_i & u_{ij} \\ \hline t_{ii} & D_1 & -E_1 \\ t_{ij} & E_2 & D_2 \end{array}$$

where $D_1 : p \times p$ and $D_2 : p(p-1)/2 \times p(p-1)/2$ are diagonal matrices with elements $u_{11}, \ldots, u_{pp}$ and x_i of multiplicity $(i-1), (i = 1, \ldots, p)$, respectively. The latter arise from $\partial t_{ij}/\partial u_{ij} = x_i\,(j = 1, 2, \ldots, i-1)$. $E_1 : p \times p(p-1)/2$ has elements $x_i u_{ij}/u_{ii}$ arising from $\partial t_{ii}/\partial u_{ij}\,(j = 1, 2, \ldots, i-1;\ i = 1, \ldots, p)$ and zeros elsewhere. $E_2 : p(p-1)/2 \times p$ has elements u_{ij} arising from $\partial t_{ij}/\partial x_i\,(j = 1, 2, \ldots, i-1\,;\ i = 1, \ldots, p)$ and zeros elsewhere. The determinant of partial derivatives is

$$|\,D_2\,|\,|\,D_1 + E_1 D_2^{-1} E_2\,| = \prod_1^p x_i^{i-1} \left| \left(\frac{u_{ii}^2 + \sum_{j=1}^{i-1} u_{ij}^2}{u_{ii}} \right) \right| = \prod_1^p x_i^{i-1} u_{ii}^{-1}.$$

THEOREM 5. The J of the transformation $\bar{R} = \tilde{U}\tilde{U}'$, where $r_{ii} = 1$, $\sum_{j=1}^{i} u_{ij}^2 = 1$, $(i = 1, \ldots, p)$ is

$$D(\tilde{U};\, \bar{R}) = \prod_1^p u_{ii}^{p-i}.$$

Proof. $r_{ii} = 1$, $r_{ij} = \sum_{k=1}^{i} u_{ki} u_{kj}\,(i < j)$. The scheme of coefficients is triangular with diagonal elements u_{ii} of multiplicity $(i-1)$ arising from $\partial r_{ij}/\partial u_{ij}\,(j = 1, \ldots, i-1)$. Hence the determinant of partial derivatives is $\prod_1^p u_{ii}^{p-i}$.

THEOREM 6. The J of the transformation $Y = (A+X)^{-1}(A-X)$, where A is a matrix of scalars $\neq 0$, is

$$D(X;\,Y) = 2^{p^2}\,|\,A\,|^{p}\,|\,A+X\,|^{-2p}.$$

Proof. Taking differentials of $Y = (A+X)^{-1}(A-X)$, we get

$$(dY) = -\,2(A+X)^{-1}\,(dX)\,(A+X)^{-1}A.$$

Using Theorem 3·6 of the earlier paper,

$$D(X;\,Y) = |\,2(A+X)^{-1}\,|^{p}\,|\,(A+X)^{-1}A\,|^{p}.$$

THEOREM 7. The J of the transformation

$$\overline{Y} = (I+\overline{X})^{-1}(I-\overline{X}) \quad \text{is} \quad D(\overline{X};\,\overline{Y}) = 2^{p(p+1)/2}\,|\,I+\overline{X}\,|^{-(p+1)}.$$

Proof. Taking differentials of $\overline{Y} = (I+\overline{X})^{-1}(I-\overline{X})$, we get

$$(d\overline{Y}) = -\,2(I+\overline{X})^{-1}\,(d\overline{X})\,(I+\overline{X})^{-1}.$$

Using Theorem 3·7 of the earlier paper,

$$D(\overline{X};\,\overline{Y}) = |\,2(I+\overline{X})^{-1}\,|^{p+1}.$$

THEOREM 8. The J of the transformation $\tilde{Y} = (\tilde{A}+\tilde{X})^{-1}(\tilde{A}-\tilde{X})$, where $\tilde{A}$ is a matrix of scalars $\neq 0$, is

$$D(\tilde{X};\,\tilde{Y}) = 2^{p(p+1)/2}\,|\,\tilde{A}+\tilde{X}\,|^{-(p+1)}\prod_{1}^{p} a_{ii}^{p-i+1}.$$

Proof. Taking differentials of $\tilde{Y} = (\tilde{A}+\tilde{X})^{-1}(\tilde{A}-\tilde{X})$, we get

$$(d\tilde{Y}) = -\,2(\tilde{A}+\tilde{X})^{-1}\,(d\tilde{X})\,(\tilde{A}+\tilde{X})^{-1}\tilde{A}.$$

Using Theorem 3·8 of the earlier paper and Theorem 1,

$$D(\tilde{X};\,\tilde{Y}) = \prod_{1}^{p} 2^{i}(a_{ii}+x_{ii})^{-i}\,(a_{ii}+x_{ii})^{-(p-i+1)}\,a_{ii}^{p-i+1}.$$

THEOREM 9. The J of the transformation $\overline{Y} = \tilde{T}D_x\tilde{T}'$, where $t_{ii} = 1$, is

$$D(\tilde{T},x\,;\,\overline{Y}) = \prod_{1}^{p} x_i^{p-i}.$$

Proof. Taking differentials of $\overline{Y} = \tilde{T}D_x\tilde{T}'$ and pre- and post-multiplying by $\tilde{T}^{-1}$ and $\tilde{T}'^{-1}$, respectively,

$$\tilde{T}^{-1}(d\overline{Y})\,\tilde{T}'^{-1} = \tilde{T}^{-1}(d\tilde{T})\,D_x + (dD_x) + D_x(d\tilde{T})'\,\tilde{T}'^{-1}.$$

Let

$$\overline{Z} = \tilde{T}^{-1}(d\overline{Y})\,\tilde{T}'^{-1}, \quad D_w = dD_x, \quad \tilde{U} = \tilde{T}(d\tilde{T}),$$

where $u_{ii} = 0$, then

$$\overline{Z} = \tilde{U}D_x + D_w + D_x\tilde{U}'$$

and

$$D(\tilde{T},x;\,\overline{Y}) = D(d\tilde{T},dx;\,d\overline{Y}) = D(d\tilde{T},dx;\,\tilde{U},dw)\,D(\tilde{U},dw;\,\overline{Z}).$$

$z_{ii} = w_{ii}$ and $z_{ij} = u_{ij}x_j$ and hence the scheme of coefficients for $D(\tilde{U},dw\,;\overline{Z})$ is

$$\begin{array}{c|cc} & w_i & u_{ij} \\ \hline z_{ii} & I & 0 \\ z_{ij} & L & D \end{array}$$

Bureau ... Ednl. ... Research
Dated ...
Accs. No ...

where $D: p(p-1)/2 \times p(p-1)/2$ is diagonal with elements x_i of multiplicity $(p-i)$ $(i = 1, \ldots, p)$. As the matrix L does not affect the determinant, we obtain

$$D(\tilde{U}, dw;\, \bar{Z}) = |\, D\,| = \prod_1^p x_i^{p-1},$$

$$D(d\tilde{T}, dx;\, \tilde{U}, dw) = |\, \tilde{T}\,|^{-(p+1)} \prod_1^p t_{ii}^{-i} = 1.$$

THEOREM 10. The J of the transformation $Y = \Gamma\tilde{T}\Gamma'$, where Γ is orthogonal such that $|\,\Gamma + I\,| \neq 0$, is

$$D(\hat{X};\, \tilde{T};\, Y) = 2^{p(p-1)/2}\, |\, I + \hat{X}\,|^{-(p-1)} \prod_{i<j} (t_{ii} - t_{jj}),$$

where

$$\Gamma = (I + \hat{X})^{-1}(I - \hat{X}).$$

Proof. Taking differentials of $Y = \Gamma\tilde{T}\Gamma'$, and pre- and post-multiplying by Γ' and Γ, respectively,

$$\Gamma'(dY)\,\Gamma = \Gamma'(d\Gamma)\,\tilde{T} + (d\tilde{T}) + \tilde{T}(d\Gamma)'\,\Gamma.$$

Let

$$W = \Gamma'(dY)\,\Gamma, \quad \hat{S} = \Gamma'(d\Gamma), \quad \tilde{U} = d\tilde{T},$$

then

(1) $$W = \hat{S}\tilde{T} + \tilde{U} - \tilde{T}\hat{S},$$

(*Note*. $\Gamma'(d\Gamma) = \hat{S}$ is skew-symmetric, hence $S'\hat{\ } = -S.)\hat{\ }$

$$D(\hat{X}, \tilde{T};\, Y) = D(d\hat{X}, dT;\, dY) = D(d\hat{X}, dT;\, \hat{S}, \tilde{U})\, D(\hat{S}, \tilde{U};\, W)\,(W;\, dY) \equiv D_1 D_2 D_3,$$

$$D_1 = 2^{p(p-1)/2}\, |\, I + \hat{X}\,|^{-(p-1)}, \quad D_3 = 1,$$

D_2 arises from (1), and the scheme of coefficients is

$$\begin{array}{l|ccc} & u_{ii} & u_{ij} & s_{ij} \\ \hline y_{ii} & I & 0 & L \\ y_{ij}\,(i<j) & 0 & I & M \\ y_{ij}\,(i<j) & 0 & 0 & N \end{array},$$

the determinant of which is $|\,N\,|$. N is a triangular matrix with diagonal elements $t_{ii} - t_{ii})$ $(I \neq j)$ arising from $\partial y_{ij}/\partial s_{ij} = t_{jj} - t_{ii}$, where

$$y_{ij} = \sum_{k=j}^{p} s_{ik} t_{jk} - \sum_{k=1}^{1} s_{kj} t_{ki}.$$

Hence

$$|\,N\,| = \prod_{i<j} (t_{ii} - t_{jj}).$$

CONCLUSION

In this note, the Jacobians of certain matrix transformations have been evaluated. Theorem 2 is concerned with an alternative form of the Toeplitz factorization which is a rotation of the rectangular co-ordinates. In Theorem 3, the relation between the correlation and covariance matrices is stated. In Theorem 4, a set of normalized rectangular co-ordinates is introduced, and in Theorem 5, these are related to the correlation matrix. In Theorems 6–8, respectively, the Cayley transformation for a general matrix, a symmetric and a triangular matrix is given. In Theorem 10, the Smith canonical form which yields the characteristic roots of a matrix is considered.

REFERENCES

DEEMER, W. L. & OLKIN, I. (1951). The Jacobians of certain matrix transformations useful in multivariate analysis, based on Lectures by P. L. Hsu. *Biometrika*, **38**, 345.

ESTIMATION OF A FUNCTIONAL RELATIONSHIP*

By D. V. LINDLEY

Statistical Laboratory, University of Cambridge

It is known (Lindley, 1947, § 7·2) that if there is a linear functional relationship between two quantities, if the only observations available on these quantities are subject to error and if all the distributions are normal, then there is no satisfactory estimate of the functional relationship unless the relative accuracy of measurement of the two quantities, or some similar knowledge, is available. Furthermore, in this situation there are three relationships which exist between the two quantities (either with or without error), namely, the original functional relationship and two regression equations which, in this case, will be linear. The experimenter who encounters this situation is therefore faced with a difficulty and a complexity in the mathematics which he feels is foreign to his practical situation, and it is not therefore surprising that numerous attempts have been made to remove the undesirable features of the problem. The fact that the experimenter finds difficulties which he cannot readily interpret suggests that the mathematical model is unsatisfactory. A new model has therefore been suggested by Berkson (1950), and with this model he claims that the estimation is possible, that there is only one regression and that this regression has the same constants as the functional relationship. Kendall (1952) has raised some doubts about Berkson's procedure, and when I read Prof. Kendall's paper before publication I thought I agreed with him. Further consideration has revealed that Berkson's procedure for estimating the functional relationship is sound and his mathematical model, incorporating the idea of a controlled observation, is likely to be of very wide applicability. In the present paper I give a mathematical justification of Berkson's procedure which may help to clarify the issue.

As a compromise between the notations of Berkson and Kendall random variables (variates in Kendall's terminology) will be denoted by lower case Roman letters and other quantities (mathematical variables) by capital letters. Any random variable or quantity which can be observed will have a prime attached to it. We now postulate the existence of two quantities, U and V, which are connected by a linear functional relationship

$$V = \alpha + \beta U. \tag{1}$$

Nothing is said about how U and V are to be obtained; they may be obtained by a random process so that (1) may equally be written

$$v = \alpha + \beta u, \tag{2}$$

but, in view of the relationship, if one is a random variable so is the other. In the problem under consideration U and V are not observable; observations are only available on

$$x' = u + d, \tag{3}$$

$$y' = v + e, \tag{4}$$

---

* See editorial note on p. 49 below.

where d and e are two random variables. The problem is to estimate α and β from several pairs of observations on x' and y'. If we now assume

(i) that u (and hence v) is normally distributed,
(ii) d and e are independently normally distributed with zero means, and
(iii) d and e are independent of u,

then estimation of α and β is not possible and the regression lines of y' on x' and x' on y' will not be coincident and neither will have the constants of the functional relationship. The estimation does not seem possible even if (i) is omitted and U is not a random variable. This is the difficulty referred to above.

Berkson queries the formulation (3) and the assumption (iii). Taken together these imply that we have a true value u (or U) which we are unable to measure directly but in making a measurement on u we introduce a random error d which is unbiased and *independent of* u so that x' is of necessity a random variable. Berkson calls x' an *uncontrolled* observation. He suggests that observations are not always uncontrolled; often we bring a measurement to a value determined beforehand (whether or not by a random process will be seen not to matter) so that our measured quantity is fixed but the actual value of the quantity used in the experiment will differ by an error d from the measurement. Hence mathematically if we take the case where the measurement is not a random variable

$$u = X' - d, \tag{5}$$

where we may assume d to be normally distributed with zero mean, and the minus sign is introduced in order to show the analogy between (3) and (6). Rewriting (5)

$$X' = u + d, \tag{6}$$

and u and d *are no longer independent*. X' is called a *controlled* observation. Berkson suggests that in many experiments one value is controlled, the other is uncontrolled; for example we may bring the current to 2 amp. (controlled) and measure the voltage as 3·67 V. (uncontrolled). In this case Berkson argues that estimation of α and β is possible and the regression of y' on X' is linear with the same constant. The proof of these statements is simple. From (2), (4) and (5)

$$y' - e = \alpha + \beta(X' - d),$$

i.e.

$$\begin{aligned} y' &= \alpha + \beta X' + (e - \beta d) \\ &= \alpha + \beta X' + f, \end{aligned} \tag{7}$$

where $f = e - \beta d$ is a random variable with zero mean and normally distributed. This is the ordinary regression of a random variable y on a fixed variable X, and accordingly the regression stated is established and the usual estimates

$$b = \frac{\Sigma(y' - \bar{y}')(X' - \bar{X}')}{\Sigma(X' - \bar{X}')^2}, \quad a = \bar{y}' - b\bar{X}',$$

of β and α are available and have the usual optimum properties. Furthermore, even if X' is a random variable (which in most applications it will not be) Berkson's argument persists because *f is independent of x'*. In the original model where we have (2), (3) and (4) we obtain the equation

$$y = \alpha + \beta x' + f, \tag{8}$$

analogous to (7) but x' and f *are correlated*, since x' is correlated with d, and thus the usual regression analysis ceases to apply. A finer analysis shows the estimation to be impossible.

Mathematically the essence of Berkson's idea is to suppose d independent of x' (where I mean statistical independence in the case where x is a random variable) in contrast with the usual assumption that d is independent of u. The above analysis shows that in Berkson's situation the problem is equivalent to that of the regression of y' on X' (or x'). If X is not a random variable then Berkson's procedure can be justified without the assumption of normality by the familiar least squares argument and an appeal to Gauss's theorem.

Since it has caused confusion in the past it might be helpful to rephrase Berkson's first conclusion (p. 173). If x' is controlled and y' is not, then the regression of y' on x' is of the usual form, has the same constants as the functional relationship, and the latter may be found by the usual regression methods. If y' is controlled and x' is not, the regression of x' on y' has similar properties. But note that in the former case there will still be a regression of x' on y' though it will not have the constants of the functional relationship; it may not even be linear. So that Berkson's title 'Are there two regressions?' is still answered in the affirmative if the controlled observation is a random variable (a case not considered by him), but in the negative when it is a mathematical variable, since regression of a mathematical variable has no meaning. Nevertheless, his procedure for estimating the functional relationship, which is quite a different thing, as Kendall points out, is valid.

REFERENCES

BERKSON, J. (1950). *J. Amer. Statist. Ass.* **45**, 164.
KENDALL, M. G. (1952). *Biometrika*, **39**, 96.
LINDLEY, D. V. (1947). *Suppl. J.R. Statist. Soc.* **9**, 218.

[*Editorial Note.* In his second paper on 'Regression, structure and functional relationship', Kendall (1952) criticized certain of Berkson's (1950) conclusions regarding regression problems when dealing with controlled experiments. To these criticisms Dr Berkson submitted a reply which in turn led to more comments from Prof. Kendall. In considering these further contributions as Editor, I recalled how often in controversy disagreement arises from quite inadvertent misunderstanding of the other man's terminology, notation or method of approach; to clear the position an independent re-statement of the problems at issue may be needed. The preceding note by Mr D. V. Lindley appeared to be so helpful in this connexion that I asked the two contestants whether they would agree to its publication in place of their own explanations. This plan they both welcomed; it is, however, right that I should add that, while Dr Berkson finds himself in substantial agreement with what Mr Lindley has written, Prof. Kendall is not convinced that the points at issue have all been cleared. E.S.P.]

ESTIMATING PARAMETERS IN TRUNCATED PEARSON FREQUENCY DISTRIBUTIONS WITHOUT RESORT TO HIGHER MOMENTS

By A. C. COHEN, Jr., *The University of Georgia*

1. Introduction

Truncated samples arise whenever observed measurements are restricted to an interval which fails to include the full range of population values. Numerous biological, psychological, sociological, economic, engineering and other scientific investigations produce such samples. They result from studies of 'time to death' or 'time to react' of toxicants, stimulants, etc., when observation is discontinued before all exposed specimens exhibit the reaction under study. They likewise arise when sorting or inspection procedures eliminate specimens above or below designated tolerance limits, as well as in many other practical situations. Until recently, interest in truncated samples has been largely limited to those from normal populations. The present paper is more general in scope and is concerned with truncated samples from all types of Pearson populations.

Without resorting to higher moments as was done in an earlier work by the writer (1951), the present paper treats the specific problem of estimating parameters of univariate Pearson populations from samples truncated at known terminal points when the number of observations thus eliminated is unknown. Using the method of moments, estimating equations are obtained which involve only the first four sample moments in contrast to the first six which were previously employed. The Pearson-Lee-Fisher estimating equations (1908, 1915, 1931) for singly truncated samples from normal populations follow as a special case of the present results. Likewise, previous results of the writer (1950*a*) pertaining to singly truncated type III distributions also constitute a special case. The practical application of the results of this paper is illustrated using a doubly truncated sample from a type III population. Throughout this paper, samples are described as singly or doubly truncated according to whether one or both 'tails' are missing.

2. Some fundamental moment relationships

In the writer's paper (1951), previously referenced, it was demonstrated that the system of moment equations

$$h\mu'_k + b_0 k\mu'_{k-1} + b_1 k\mu'_k + b_2(k+2)\,\mu'_{k+1} = \mu'_{k+1} \quad (k = 0, 1, 2, 3), \tag{1}$$

where $h = a + b_1$ and μ'_k is the kth moment of the general Pearson frequency function, follows readily from the differential equation

$$\frac{1}{f(x)}\frac{df(x)}{dx} = \frac{a-x}{b_0 + b_1 x + b_2 x^2}. \tag{2}$$

Equations (1) differ from the usual equations for the Pearson system only in the use of h in place of a. They were, of course, derived from (2) on the assumption that the expression $(b_0 + b_1 x + b_2 x^2)\, x^k f(x)$ vanishes at both terminals of $f(x)$. For the Pearson curves of types IV, VI and VII having a range unlimited in one or both directions, this will not be true when k exceeds a certain value which depends on b_2.

If the origin from which x is measured is designated ξ in standard units referred to the distribution mean, then

$$\mu_1' = -\sigma\xi, \tag{3}$$

where σ is the standard deviation.

If now the non-central moments of (1) are replaced by standard moments α_k,† where

$$\alpha_k = \sum_{i=0}^{k} \binom{k}{i} \mu_{k-i}'(-\mu_1')^i/\sigma^k, \tag{4}$$

the resulting system can be solved to yield

$$\left.\begin{aligned} h = \sigma\lambda, \quad b_0 = \sigma^2\theta, \quad b_1 = \sigma\psi, \\ b_2 = \frac{6+3\alpha_3^2-2\alpha_4}{2(9+6\alpha_3^2-5\alpha_4)}, \end{aligned}\right\} \tag{5}$$

where

$$\left.\begin{aligned} \lambda &= -\xi(1-2b_2), \\ \theta &= [(1-3b_2)+\xi\{(1-4b_2)\alpha_3/2+\xi b_2\}], \\ \psi &= [\alpha_3/2-2b_2(\alpha_3-\xi)]. \end{aligned}\right\} \tag{6}$$

3. GENERAL FOUR-PARAMETER DISTRIBUTIONS

Doubly-truncated samples. If $f(x)$ is truncated both on the left and on the right, and if the origin is taken at the left terminus so that $0 \leqslant x \leqslant d$, where d is thus the truncated range, the following system of equations may be obtained in a manner similar to that by which (1) was derived:

$$\left.\begin{aligned} &h+2m_1b_2+F_1-F = m_1, \\ &hm_k+b_0km_{k-1}+b_1m_k+b_2(k+2)\,m_{k+1}-d^kF = m_{k+1} \quad (k=1,2,3), \end{aligned}\right\} \tag{7}$$

where

$$F_1 = f(0)\,b_0, \quad F = f(d)\,[b_0+db_1+d^2b_2], \tag{8}$$

and m_k is the kth moment of the truncated distribution about its left terminus, i.e.

$$m_k = \int_0^d x^k f(x)\,dx \quad \text{with } m_0 = 1. \tag{9}$$

It should be noted that truncation does not alter the differential equation. The constants b in (7) are therefore identical with those in (2), and may be expressed through (5) and (6) in terms of moments of the complete distribution. Further details are included in the writer's earlier paper.

The left and right truncation points in standard units of the complete distribution referred to its mean are designated as ξ' and ξ'', respectively, with $\xi'' = \xi' + d/\sigma$. We can write

$$F_1 = \sigma Z_1 \quad \text{and} \quad F = \sigma Z_2, \tag{10}$$

where

$$\left.\begin{aligned} Z_1 &= [\phi(\xi')/p]\,[(1-3b_2)+\xi'\{(1-4b_2)\,\alpha_3/2+\xi'b_2\}], \\ Z_2 &= [\phi(\xi'')/p]\Big\{[(1-3b_2)+\xi'\{(1-4b_2)\,\alpha_3/2+\xi'b_2\}]+(d/\sigma)\,[\alpha_3/2-2b_2(\alpha_3-\xi')]+(d/\sigma)^2\,b_2\Big\}. \end{aligned}\right\} \tag{11}$$

---

† In the original Pearson notation, $\alpha_3 = \sqrt{\beta_1}$, $\alpha_4 = \beta_2$.

The foregoing results follow from (8) when we substitute the values given in (5) and (6) for b_0 and b_1 and simultaneously substitute for $f(0)$ and $f(d)$ the ratio of standardized ordinate to area

$$f(0) = \phi(\xi')/(p\sigma) \quad \text{and} \quad f(d) = \phi(\xi'')/(p\sigma), \quad \text{where } p = \int_{\xi'}^{\xi''} \phi(t)\,dt. \tag{12}$$

If the values given in (10) and (11) for F_1 and F and those in (5) for h, b_1 and b_0 are substituted into (7), the resulting system of equations becomes

$$\left.\begin{aligned} m_1 &= \sigma[Z_1 - Z_2 + \lambda]/(1-2b_2) &&= \sigma g_1,\\ m_2 &= \sigma^2[(\lambda+\psi)\,g_1 + \theta - (d/\sigma)\,Z_2]/(1-3b_2) &&= \sigma^2 g_2,\\ m_3 &= \sigma^3[(\lambda+2\psi)\,g_2 + 2\theta g_1 - (d/\sigma)^2\,Z_2]/(1-4b_2) &&= \sigma^3 g_3,\\ m_4 &= \sigma^4[(\lambda+3\psi)\,g_3 + 3\theta g_2 - (d/\sigma)^3\,Z_2]/(1-5b_2) &&= \sigma^4 g_4,\end{aligned}\right\} \tag{13}$$

where g_1, g_2, g_3 and g_4 are functions of ξ', σ, α_3 and α_4 as defined above. It must be remembered that θ, λ and ψ as defined in (6) are functions of ξ', α_3 and α_4 but not of σ.

When observed sample moments are equated to corresponding population moments, i.e. $\nu_k = m_k$, where ν_k is the kth observed moment of the truncated sample about its left terminus $\left(\nu_k = \sum_{i=1}^{n} x_i^k/n\right)$, and these values are substituted into (13), we obtain the system of estimating equations

$$\nu_r = \sigma^{*r} g_r(\xi'^*, \sigma^*, \alpha_3^*, \alpha_4^*) \quad (r = 1, 2, 3, 4). \tag{14}$$

The stars (*) were added to distinguish estimates from parameters. After solving (14) for estimates, $\xi'^*, \sigma^*, \alpha_3^*$ and α_4^*, the mean can be estimated with the aid of (3) as

$$\mu_1'^* = -\sigma^*\xi'^*. \tag{15}$$

In practice, solution of the complete system (14), which applies to general four-parameter distributions when samples are doubly truncated, is likely to be rather tedious. These equations can, however, be solved by an iterative process similar to one described by Scarborough (1930). First approximations might be obtained by using the writer's method of higher moments (1951) or perhaps even by judicious guessing. If the truncation is not too severe, the usual moment estimates computed using complete sample formulae might suffice.

Singly-truncated samples. When only a single 'tail' is missing from the sample, g_1, g_2, g_3 and g_4 are functions of ξ', α_3 and α_4 only. Accordingly, by eliminating σ between them, the four equations of (14) are in this instance reduced to three in number. The labour involved in their solution is thereby greatly reduced.

4. TYPE III AND NORMAL DISTRIBUTIONS

Type III distributions. When $f(x)$ is of type III, then $b_2 = 0$, and equations (5) and (6) assume the more simple forms

$$\left.\begin{aligned} h &= -\sigma\xi, & \lambda &= -\xi,\\ b_0 &= \sigma^2(1+\xi\alpha_3/2), & \theta &= 1+\xi\alpha_3/2,\\ b_1 &= \sigma\alpha_3/2, & \psi &= \alpha_3/2.\end{aligned}\right\} \tag{16}$$

Subsequently, it follows from the above results together with equations (13) that for *doubly-truncated type III* samples

$$\left.\begin{aligned} g_1 &= [Z_1 - Z_2 - \xi'],\\ g_2 &= [(\alpha_3/2 - \xi')\,g_1 + (1 + \xi'\alpha_3/2) - (d/\sigma)\,Z_2],\\ g_3 &= [(\alpha_3 - \xi')\,g_2 + 2(1 + \xi'\alpha_3/2)\,g_1 - (d/\sigma)^2\,Z_2].\end{aligned}\right\} \tag{17}$$

Similarly, it follows that

$$\left.\begin{aligned} Z_1 &= [\phi(\xi')/p]\,[1 + \xi'\alpha_3/2],\\ Z_2 &= [\phi(\xi'')/p]\,[(1 + \xi'\alpha_3/2) + (d/\sigma)\,(\alpha_3/2)].\end{aligned}\right\} \tag{18}$$

Since only the three estimates ξ'^*, σ^* and α_3^* are required in this case, it is necessary only to let $r = 1, 2$ and 3 in (14) and then solve the three resulting equations.

With *singly-truncated type III* samples, the first three equations of (14) immediately reduce to two when σ is eliminated between them. This case has been considered separately (1950*a*), and no further comment seems necessary here except to note that the function Z of the present paper was given as $Z\theta$ in the earlier work.

Truncated normal distributions. When $f(x)$ is the normal distribution, $b_1 = b_2 = 0$, and the estimating equations of (14) reduce to identical forms obtained previously using the method of maximum likelihood for both singly- and doubly-truncated samples (1949, 1950*b*).

5. A PRACTICAL APPLICATION

The practical application of results of this paper can be conveniently demonstrated with the same example previously employed to illustrate the method of higher moments. The original data consist of weights of 1000 female students (cf. Table 1) as given by Miss Shook (1930). Miss Shook computed estimates of population parameters from this sample under the assumption that it was a complete sample from a type III population. For the present illustration, a doubly-truncated sample was obtained from Miss Shook's data by arbitrary truncation at 79·95 and 159·95 lb. respectively. Thereby the first and the last six cells of the grouped data were eliminated, and the truncated sample thus formed consists of 981 observations within the interval bounded by the above terminal points.

The first three sample moments about the left truncation point both before and after making appropriate corrections for grouping errors are given below:

Uncorrected moments	Corrected moments
$n_1 = 37{\cdot}833843$	$\nu_1 = 37{\cdot}833843$
$n_2 = 1660{\cdot}06626$	$\nu_2 = 1651{\cdot}73293$
$n_3 = 81{,}227{\cdot}1916$	$\nu_3 = 80{,}281{\cdot}3455$

Grouping corrections which hold only when high contact conditions at the terminal points are satisfied are not applicable. Those employed here are due to Craig (1937), which he describes as Sheppard's corrections for discrete data, derived under conditions which do not demand high contact. In justification of their use in the present instance, note that Miss Shook's original raw data may be viewed as consisting of discrete measurements, each recorded to the nearest $\frac{1}{10}$ lb. with grouping accomplished by placing sequences of 100 consecutive values of these discrete values into successive 10 lb. intervals. When final graduations are compared with observed data, the foregoing contentions seem to be further substantiated.

Since our sample is doubly truncated from a type III population, the required estimates can be computed from the first three equations of (14). In the present instance, however, it seems preferable as a matter of practical convenience to turn back to the first three equations of (7) with $b_2 = 0$. From these we compute estimates of h, b_0 and b_1. Subsequently we calculate estimates of ξ', σ and α_3 using the relations

$$\sigma = \sqrt{(b_0+b_1h)}, \quad \alpha_3 = 2b_1/\sigma \quad \text{and} \quad \xi' = -h/\sigma,$$

which follow from equation (16).

On equating observed sample moments to corresponding population moments, estimating equations for the truncated sample under consideration become

$$\begin{aligned} h^* &= 37{\cdot}833843 - F_1^* + F^*, \\ b_0^* + 37{\cdot}833843 b_1^* &= 1651{\cdot}73293 + 80F^* - 37{\cdot}833843h^*, \\ 75{\cdot}667686 b_0^* + 3303{\cdot}46586 b_1^* &= 80281{\cdot}3455 + 6400F^* - 1651{\cdot}73293h^*. \end{aligned}$$

Starting with first approximations $\xi_1' = -2{\cdot}5$, $\sigma_1 = 15$ and $\alpha_{3_1} = 0{\cdot}35$, which were calculated as usual moment estimates, neglecting the fact that our sample is actually truncated, we employ the iterative method described by Scarborough (1930) to improve these results. After seven iterations,† we obtain $\xi'^* = -2{\cdot}43$, $\sigma^* = 15{\cdot}85$ and $\alpha_3^* = 0{\cdot}61$ which are correct to the number of significant digits given. Using these values, it follows from (15) that $\mu_1'^* = 38{\cdot}48$ and subsequently $M^* = 79{\cdot}95 + \mu_1'^* = 118{\cdot}43$ lb. (mean measured from 0).

Although convergence of the iterative process in this instance is not as rapid as might be desired, the number of iterations required was increased somewhat as a consequence of the rather wide error in the first approximation to α_3 (0·35). When calculations were repeated using as first approximations, $\alpha_{3_1} = 0{\cdot}66$, $\sigma_1 = 16$ and $\xi_1' = -2{\cdot}42$ which were obtained as higher moment estimates of population parameters, the iterative process led to the same final estimates as before, after only two iterations instead of seven.

A graduation of the sample data based on the estimates obtained here was performed with the aid of Salvosa's Tables (1930), and these results along with Miss Shook's original graduation based on estimates computed from the complete sample are included in Table 1. Also included in this table is a graduation taken from the writer's earlier paper (1951) which was based on higher moment estimates from the same doubly-truncated sample considered here.

It is observed that graduated frequencies based on estimates computed from the truncated sample are in closer agreement with observed frequencies than those based on complete sample moment estimates. This may be due largely to the fact that complete sample moment estimates are unduly affected by extreme observations which are eliminated from the truncated sample. At least it has been the writer's experience to date that moment estimates of long-tailed Pearson distributions are improved by truncating the sample tails, and subsequently calculating estimates either by the method of this paper or by the higher moment method of the earlier paper. Although the graduation based on estimates computed from the first three truncated sample moments by the method of this paper does not appear to be appreciably better than the higher moment graduation based on the first five moments, the present estimates would be expected to exhibit smaller variances as a consequence of not using the higher order moments in their calculation. Either set of estimates, for the present illustration, would certainly be considered adequate as first approximations on which to base iterations to maximum-likelihood estimates. This topic is one which the writer hopes to investigate more fully in a subsequent paper.

---

† An illustration of one of these iterative steps is given in the final paragraph of this section.

Table 1. *Weights of* 1000 *female students*

Weight (lb.)	Observed frequency	Graduated frequencies		
			Doubly truncated sample	
		Complete sample (Miss Shook's graduation)	Graduation based on higher sample moments (first five moments)	Graduation based on estimates of this paper (first three moments)
70–79·9	2	0	0·2	0·2
80–89·9	16	4	12·7	13·8
90–99·9	82	102	94·0	93·2
100–109·9	231	238	214·0	212·1
110–119·9	248	250	253·9	255·0
120–129·9	196	184	200·8	202·9
130–139·9	122	111	120·6	121·2
140–149·9	63	59	59·4	58·5
150–159·9	23	29	25·2	24·1
160–169·9	5	13	9·5	8·7
170–179·9	7	6	3·3	2·9
180–189·9	1	3	1·0	0·9
190–199·9	2	1	0·3	0·2
200–209·9	1	0	0·1	0·1
210–219·9	1	0	0·0	0·0
Total	1000	1000	995·0	993·6
Total frequency in truncated range	981	977	980·6	980·8
M^* (lb.)	—	118·74	118·56	118·43
σ^* (lb.)	—	16·9175	16·02	15·85
α_3^*	—	0·976424	0·657	0·61
Lower limit (lb.)	—	84·09	69·77	66·46

Truncated sample formed by truncating the complete sample on the left at 79·95 lb., and on the right at 159·95 lb.

A sample of the calculations involved in one cycle of the iterative process will serve to illustrate details of the technique. At the end of two cycles, we obtain as third approximations $(\alpha_3)_3 = 0{\cdot}50$, $\sigma_3 = 15{\cdot}5$ and $\xi_3' = -2{\cdot}44$. Since $\xi'' = \xi' + d/\sigma$, and since for the example under consideration, $d = 80$, it follows that $\xi_3'' = -2{\cdot}44 + 80/15{\cdot}5 = 2{\cdot}72$. From Salvosa's tables, we read

$$\phi(-2{\cdot}44) = 0{\cdot}005050, \quad \phi(2{\cdot}72) = 0{\cdot}017913 \quad \text{and} \quad p = \int_{-2\cdot 44}^{2\cdot 72} \phi(t)\,dt = 0{\cdot}989868.$$

On substituting the above values in equations (18) we compute:

$$Z_1 = \left[\frac{0{\cdot}005050}{0{\cdot}989868}\right][1 + (-2{\cdot}44)(0{\cdot}5)/2] = 0{\cdot}00199,$$

and
$$Z_2 = \left[\frac{0{\cdot}017913}{0{\cdot}989868}\right][1 + (-2{\cdot}44)(0{\cdot}5)/2 + (80/15{\cdot}5)(0{\cdot}5/2)] = 0{\cdot}0304.$$

Finally, from equation (10) we have as third approximations,

$$F_1 = 15{\cdot}5\,(0{\cdot}00199) = 0{\cdot}03 \quad \text{and} \quad F = 15{\cdot}5\,(0{\cdot}0304) = 0{\cdot}47.$$

These values are rounded off to 0 and 0·5 respectively, and the estimating equations for the present cycle of calculations then become

$$\begin{aligned} h &= 38{\cdot}334, \\ b_0 + 37{\cdot}834 b_1 &= 241{\cdot}416, \\ b_0 + 43{\cdot}658 b_1 &= 266{\cdot}482. \end{aligned}$$

On solving for b_0 and b_1 we obtain $b_1 = 4{\cdot}304$ and $b_0 = 78{\cdot}577$. Subsequently we calculate fourth approximations

$$\begin{aligned} \sigma_4 &= \sqrt{\{78{\cdot}577 + (4{\cdot}304)(38{\cdot}334)\}} = 15{\cdot}6, \\ (\alpha_3)_4 &= 2(4{\cdot}304)/15{\cdot}6 = 0{\cdot}55, \\ \xi_4' &= -38{\cdot}334/15{\cdot}6 = -2{\cdot}43. \end{aligned}$$

Using these values as new (fourth) approximations, we repeat the process as required until the desired accuracy is achieved. As stated previously, seven iterations were necessary to compute the final estimates given for the present example. In the subsequent iterations, more decimals were carried and the rounding off of F_1 and F was not quite so drastic as in the cycle illustrated above.

6. CONCLUDING REMARKS

Computations for applying the method of this paper to practical problems can be carried out with the aid only of ordinary tables of areas and ordinates of appropriate Pearson curves. However, for any routine applications it would be extremely helpful to have tables of F_1 and F and possibly also of g_1, g_2, g_3 and g_4 for the principal Pearson distributions.

Asymptotic variances of estimates given in this paper, although somewhat unwieldy in form, are of the order of $(1/n)$ and can be approximated by the so-called delta method as described by Cramér (1946). Further investigation of the small sample variances of estimates of this paper remains to be undertaken.

REFERENCES

COHEN, A. C. JR. (1949). On estimating the mean and standard deviation of truncated normal distributions. *J. Amer. Statist. Ass.* **44**, 518.

COHEN, A. C. JR. (1950*a*). Estimating parameters of Pearson type III populations from truncated samples. *J. Amer. Statist. Ass.* **45**, 411.

COHEN, A. C. JR. (1950*b*). Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples. *Ann. Math. Statist.* **21**, 557.

COHEN, A. C. JR. (1951). Estimation of parameters in truncated Pearson frequency distributions. *Ann. Math. Statist.* **22**, 256.

CRAIG, C. C. (1937). Sheppard's corrections for a discrete variable. *Ann. Math. Statist.* **7**, 55.

CRAMÉR, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press.

FISHER, R. A. (1931). Properties and applications of *Hh* functions. *Introduction to Mathematical Tables*, **1**, xxvi–xxxv. *Brit. Ass. Adv. Sci.*

LEE, ALICE (1915). Table of Gaussian 'tail' functions when the 'tail' is larger than the body. *Biometrika*, **10**, 208.

PEARSON, KARL & LEE, ALICE (1908). On the generalized probable error in multiple normal correlation. *Biometrika*, **6**, 59.

SALVOSA, LUIS R. (1930). Tables of Pearson's type III function. *Ann. Math. Statist.* **1**, appended.

SCARBOROUGH, J. B. (1930). *Numerical Analysis*, pp. 191–3. Johns Hopkins Press, Baltimore, Md.

SHOOK, B. L. (1930). Synopsis of elementary mathematical statistics. *Ann. Math. Statist.* **1**, 14.

A PROBLEM OF INTERFERENCE BETWEEN TWO QUEUES

By J. C. TANNER

Road Research Laboratory, Department of Scientific and Industrial Research

1. Introduction

In most of the queueing situations which are amenable to mathematical analysis, the queueing units (customers in a shop, etc.) are all trying to achieve the same object, for example, in the case of a shop, to be served. Usually, the only differences between successive customers are their times of arrival, and perhaps the time they take to be served. Congestion then arises because only a limited number of customers can be served at a time. A feature of many road-traffic queueing situations, however, is that there are a number of different 'populations' of 'customers' (in this case vehicles), and congestion arises not so much because too many vehicles want to do the same thing at the same time, but because a number of vehicles want to do different things which cannot all be done simultaneously. For example, at a cross-roads, less congestion is usually caused by a number of vehicles proceeding in the same direction on the same road than by a smaller number of vehicles, half on one road and half on another.

The problem to be considered here is one of the simpler ones of the above type. It concerns the delays that occur when two opposing streams of vehicles are trying to pass along a length of road only wide enough for one vehicle at a time. The mathematical difficulties arising from the necessity of considering the behaviour of two queues simultaneously are, not surprisingly, rather great, and the problem as first formulated has not been completely solved. Nevertheless, the results obtained have applications to several other road-traffic problems and these are briefly discussed. There may also be applications to other situations involving the flow of discrete objects, for example, articles produced in a factory.

2. Statement of the problem

The physical situation associated with the statement and analysis of the problem has been chosen as one which is fairly easy to visualize. It is not necessarily the most useful application of the results.

We have a road of which a certain length AB is only wide enough for a single traffic lane and outside which there is free flow in both directions. Vehicles, referred to as V_1's, arrive at A at random at an average rate q_1 per unit time, wishing to pass from A to B. Similarly, vehicles V_2 arrive at B at random at an average rate q_2 per unit time, wishing to pass from B to A. A vehicle V_1 will pass straight into AB on arrival at A if there is no opposing vehicle V_2 in AB, and if no other V_1 has entered AB during the previous time β_1. Otherwise it will wait until both these conditions are satisfied. Similarly, a V_2 will pass into AB at B as soon as there is no V_1 in AB and no other V_2 has entered AB within the previous time β_2. We further assume that each V_1 takes exactly a time α_1 to pass through AB, and that a V_2 takes exactly α_2. It is assumed that all vehicles travel at constant speed and that starting and stopping times are negligible, though these can to some extent be allowed for by adjusting the constants α_1 and α_2.

The problem is to find the average delay $\overline{w}_1$ to vehicles V_1 and $\overline{w}_2$ to vehicles V_2 when a stationary solution exists, i.e. when the rate at which vehicles can pass through AB is greater than the rate at which, in the long run, they will arrive, so that queues of vehicles will always clear themselves, given sufficient time. Other aspects of the process might be of interest, for example, the frequency distribution of the waiting times, but the average delays are considered to be more important and are easier to find.

There are three distinct cases which should be considered, according to the relative magnitudes of the α's and β's:

$$\text{(i)} \quad \alpha_1 > \beta_1, \ \alpha_2 > \beta_2,$$

$$\text{(ii)} \quad \alpha_1 > \beta_1, \ \alpha_2 < \beta_2 \quad (\text{or } \alpha_1 < \beta_1, \ \alpha_2 > \beta_2),$$

$$\text{(iii)} \quad \alpha_1 < \beta_1, \ \alpha_2 < \beta_2.$$

Cases (ii) and (iii) appear to have fewer applications and present mathematical difficulties of rather greater complexity than case (i). We therefore confine attention to the latter case.

The physical situation associated with case (i) is as follows: If there is a V_1 in AB, then V_2's will not have an opportunity to enter AB until there are no V_1's in AB or waiting at A, since there will, until then, always be at least one V_1 in AB. Similar considerations apply for the other stream, and we may say that at any moment either V_1's or V_2's or no vehicles have control of AB, according to whether it is occupied by V_1's, V_2's or is empty. In cases (ii) and (iii), a queue of vehicles may not be able to clear in a single 'block'; vehicles in the other stream may be able to interrupt the flow, even though the vehicles in the first are at their minimum separation.

The concept of regeneration points (Kendall, 1951, p. 184) gives an indication of why cases (ii) and (iii) appear to be less amenable to mathematical treatment. In case (i) there are various convenient sets of regeneration points; for example, the beginnings of all blocks of V_1's. In case (ii), the only comparable set appears to be the beginnings of all blocks of V_2's, while in case (iii) there are no comparable sets. The method to be used for case (i) makes use of two sets of regeneration points and so is inapplicable to cases (ii) and (iii). Of course, any defined set of instants forms a set of regeneration points in any of the three cases if the precise position of every vehicle in the system is included in the available description of the state of the system at any time, but such a set would not be of practical use.

3. ANALYSIS OF CASE (i) ($\alpha_1 > \beta_1, \alpha_2 > \beta_2$)

We carry out the analysis in the most general form, though the equations which arise have only been solved when $\beta_1 = 0$ or $\beta_2 = 0$.

At any time, as just mentioned, AB is occupied by either a V_1 or a V_2 or is empty. Periods when these three possibilities hold will thus alternate in a manner which must be determined. We then find the mean delays in terms of the distributions of the lengths of these periods.

3·1. *Notation and definitions*

A VB_1 (V_1 block) is a period when AB is controlled by V_1's, which is immediately preceded and followed by periods when this is not the case. VB_2's and VB_0's are similarly defined, the latter being periods when AB is empty.

t_0, t_1 and t_2 are lengths of VB_0's, VB_1's and VB_2's. Each t_i $(i = 0, 1, 2)$ has a distribution whose moment generating function (m.g.f.) $E(e^{t_iT})$ with respect to an argument T is $M_i(T)$.

r_1 and r_2 are the numbers of V_1's and V_2's waiting at the beginning of VB_1's and VB_2's respectively.

$t_1 \mid r_1$, with m.g.f. $M_1(T \mid r_1)$, is the length of a VB_1 conditional upon the number of V_1's being r_1 at its beginning. $t_2 \mid r_2$ is defined similarly.

ϕ_0, ϕ_1 and ϕ_2 are the relative frequencies of VB_0's, VB_1's and VB_2's, with $\phi_0+\phi_1+\phi_2 = 1$.

q_{ij} $(i \neq j)$ is the probability that a VB_j chosen at random was preceded by a VB_i.

Other symbols:

$$M_1(-q_2) = M_1, \qquad M_2(-q_1) = M_2,$$

$$\delta_1 = q_1+q_2-q_1M_1, \qquad \delta_2 = q_1+q_2-q_2M_2,$$

$$\gamma = M_1+M_2-M_1M_2,$$

$$\epsilon_1 = [e^{(\alpha_1-\beta_1)q_1}-1]/q_1, \qquad \epsilon_2 = [e^{(\alpha_2-\beta_2)q_2}-1]/q_2.$$

ν_1 is a function of T, the smaller real root of

$$\nu_1 = e^{\beta_1q_1(\nu_1-1)+\beta_1T}.$$

Similarly

$$\nu_2 = e^{\beta_2q_2(\nu_2-1)+\beta_2T},$$

$$\rho_1 = \beta_1q_1, \quad \rho_2 = \beta_2q_2.$$

It may be shown that the condition for a stationary solution to the problem to exist is $\rho_1+\rho_2 < 1$, and it is assumed throughout that when this is satisfied, all the series which arise converge and the m.g.f.'s of all the distributions exist. Proofs of these assumptions could be given, but they are thought to be sufficiently intuitive.

3·2. *Outline of method*

Stage 1: Find the q_{ij} and the ϕ_i in terms of M_1 and M_2 (and the basic constants).

Stage 2: Find the distributions of r_1 and r_2, also in terms of M_1 and M_2, by considering the block immediately preceding.

Stage 3: Find the distributions of t_1 and t_2, first for fixed r_1 and r_2, and then averaged over the distributions of these just obtained.

The assumption that the process is stationary leads to a pair of equations involving $M_1(T)$ and $M_2(T)$ which contain the information necessary to obtain these functions.

Stage 4: Find the mean delays as functions of the distributions of t_1 and t_2.

3·3. *Stage* 1

Define p_{ij} as the probability that a random VB_i is followed by a VB_j. Suppose we have a VB_i of length t_1; then the probability that it is followed by a VB_0 is the chance that no V_2 arrived during the interval t_1, i.e. it is $e^{-q_2t_1}$. Thus $p_{10} = E(e^{-q_2t_1})$, taken over the distribution of t_1. Hence

$$p_{10} = M_1.$$

Similarly

$$p_{20} = M_2, \quad p_{12} = 1-M_1, \quad p_{21} = 1-M_2.$$

Also

$$p_{01} = \frac{q_1}{q_1+q_2}, \quad p_{02} = \frac{q_2}{q_1+q_2}.$$

It is easily seen that
$$\begin{aligned}\phi_0 &= p_{10}\phi_1+p_{20}\phi_2 = q_{01}\phi_1+q_{02}\phi_2,\\ \phi_1 &= p_{21}\phi_2+p_{01}\phi_0 = q_{12}\phi_2+q_{10}\phi_0,\\ \phi_2 &= p_{02}\phi_0+p_{12}\phi_1 = q_{20}\phi_0+q_{21}\phi_1,\end{aligned}$$
and these equations, together with $q_{01}+q_{21} = 1$, etc., give
$$\left.\begin{aligned}\phi_0 &= \gamma(q_1+q_2)/[\delta_1+\delta_2+\gamma(q_1+q_2)],\\ \phi_1 &= \delta_2/[\delta_1+\delta_2+\gamma(q_1+q_2)],\\ \phi_2 &= \delta_1/[\delta_1+\delta_2+\gamma(q_1+q_2)],\end{aligned}\right\} \tag{1}$$
$$\left.\begin{aligned}q_{01} &= q_1\gamma/\delta_2, & q_{21} &= (1-M_2)\,\delta_1/\delta_2,\\ q_{02} &= q_2\gamma/\delta_1, & q_{12} &= (1-M_1)\,\delta_2/\delta_1,\\ q_{10} &= M_1\delta_2/\{\gamma(q_1+q_2)\}, & q_{20} &= M_2\delta_1/\{\gamma(q_1+q_2)\}.\end{aligned}\right\} \tag{2}$$

3·4. *Stage* 2

We consider the distribution of r_1. The chance that a random VB_1 is preceded by a VB_0 is q_{01}, and in this case $r_1 = 1$. Consider now a random VB_2 of length t_2. The chance that just r V_1's arrive during this VB_2 is
$$e^{-q_1t_2}(q_1t_2)^r/r! \quad (r = 0, 1, 2, \ldots).$$
Averaged over all VB_2's this is $\qquad E[e^{-q_1t_2}(q_1t_2)^r]/r!,$
where the expectation is taken over the distribution of t_2. Hence, considering only those VB_2's which are followed by VB_1's, the probability of just r is
$$\frac{1}{r!}E[e^{-q_1t_2}(q_1t_2)^r]/\{1-E[e^{-q_1t_2}]\} \quad (r = 1, 2, 3, \ldots).$$
The m.g.f. of this probability distribution is
$$\begin{aligned}M_r(T) &= E[e^{-q_1t_2}\sum_{r=1}^{\infty}(q_1t_2e^T)^r/r!]/(1-M_2)\\ &= E[e^{-q_1t_2}(e^{q_1t_2e^T}-1)]/(1-M_2)\\ &= [M_2(q_1(e^T-1))-M_2]/(1-M_2).\end{aligned}$$
We thus have the m.g.f. of the distribution of r_1 as
$$q_{01}e^T+q_{21}[M_2(q_1(e^T-1))-M_2]/(1-M_2) = q_1\gamma e^T/\delta_2+\delta_1[M_2(q_1(e^T-1))-M_2]/\delta_2, \tag{3}$$
and similarly that of r_2 as
$$q_2\gamma e^T/\delta_1+\delta_2[M_1(q_2(e^T-1))-M_1]/\delta_1. \tag{4}$$

3·5. *Stage* 3

We require the following lemma:

Lemma. We have a queue in which the units arrive at random, at a rate q per unit time, each unit taking a time β to be served. Initially there are r units in the queue, and the first one starts to be served. Then the m.g.f. of $n\beta$, where n is the number of units which will be served before the next time there are none being served, is ν^r where ν is the smaller real root of
$$\nu = e^{\beta q(\nu-1)+\beta T}. \tag{5}$$

When $r = 1$, the problem is that of finding the distribution of the number of individuals in all generations arising from a single ancestor, to use population mathematics terminology, and has been studied by several writers (e.g. Harris, 1948; Good, 1948; Kendall, 1951).

A proof will not be given here. The result for a general value of r follows immediately since n can be regarded as the sum of r independent components, the descendants of the r initial units.

Borel (1942) has given the frequency distribution of n for the case $r = 1$. It is of interest to note, although we do not need it here, that his method applies almost unchanged for a general value of r, the probability p_n of n individuals being

$$p_n = e^{-n\beta q}(\beta q)^{n-r}\, r\,.\,n^{n-r-1}/(n-r)! \quad (n = r,\, r+1,\, \ldots). \tag{6}$$

The m.g.f. of $n\beta$ can quite easily be found from the above expression for p_n.

We now apply this lemma to prove

$$\text{(i)}\quad M_1(T \mid r_1) = \frac{\nu_1^{r_1} e^{(\alpha_1-\beta_1)(T-q_1)}\,(q_1-T)}{q_1-T-\nu_1 q_1+\nu_1 q_1 e^{(\alpha_1-\beta_1)(T-q_1)}}, \tag{7}$$

$$\text{(ii)}\quad M_2(T \mid r_2) = \frac{\nu^{r_2} e^{(\alpha_2-\beta_2)(T-q_2)}\,(q_2-T)}{q_2-T-\nu_2 q_2+\nu_2 q_2 e^{(\alpha_2-\beta_2)(T-q_2)}}. \tag{8}$$

We consider (i). $t_1 \mid r_1$ is the time that elapses from the initial state in which r_1 V_1's are waiting till there are no V_1's in AB. This time can be split up into the following components:

(*a*) an 'r_1 wave';*

(*b*) an equal number, λ, of unit waves, and gaps whose distributions are negative-exponential between 0 and $\alpha_1-\beta_1$; λ is a random variable;

(*c*) a final time $\alpha_1-\beta_1$.

These components are distributed independently of each other, so we determine the m.g.f.'s of each and take their product to give $M_1(T \mid r_1)$.

For (*a*), the m.g.f. is $\nu_1^{r_1}$, by the lemma.

The gaps in (*b*) have the distribution

$$q_1 e^{-q_1 t}dt \Big/ \int_0^{\alpha_1-\beta_1} q_1 e^{-q_1 t}dt \quad (0 < t < \alpha_1-\beta_1),$$

whence their m.g.f. is
$$\frac{q_1}{q_1-T}\,\frac{e^{(\alpha_1-\beta_1)(T-q_1)}-1}{e^{-(\alpha_1-\beta_1)q_1}-1}.$$

The 'waves' in (*b*) have m.g.f. ν_1, by the lemma. The probability $P(\lambda)$ that we have just λ gaps and waves is the probability that the first λ gaps following waves are all less than $\alpha_1-\beta_1$ and that the next one is greater than $\alpha_1-\beta_1$. Thus

$$P(\lambda) = [1-e^{-(\alpha_1-\beta_1)q_1}]^{\lambda} e^{-(\alpha_1-\beta_1)q_1},$$

since
$$\int_0^{\alpha_1-\beta_1} q_1 e^{-q_1 t}dt = 1-e^{-(\alpha_1-\beta_1)q_1} \quad (\lambda = 0, 1, \ldots).$$

The lengths of the gaps and waves are all independent of each other and of the number of them, and so their sum has m.g.f.

$$\sum_{\lambda=0}^{\infty}[1-e^{-(\alpha_1-\beta_1)q_1}]^{\lambda} e^{-(\alpha_1-\beta_1)q_1}\,\nu_1^{\lambda}\left[\frac{q_1}{q_1-T}\,\frac{e^{(\alpha_1-\beta_1)(T-q_1)}-1}{e^{-(\alpha_1-\beta_1)q_1}-1}\right]^{\lambda} = \frac{e^{-(\alpha_1-\beta_1)q_1}}{1+\nu_1\dfrac{q_1}{q_1-T}[e^{(\alpha_1-\beta_1)(T-q_1)}-1]}.$$

Hence we finally have

$$M_1(T \mid r_1) = \frac{\nu_1^{r_1} e^{(\alpha_1-\beta_1)(T-q_1)}\,(q_1-T)}{q_1-T-\nu_1 q_1+\nu_1 q_1 e^{(\alpha_1-\beta_1)(T-q_1)}}.$$

The expression for $M_2(T \mid r_2)$ may be derived likewise.

---

* This is defined as the $n\beta$ of the lemma, and is β greater than the time that elapses until there are no V_1's waiting.

We next average $M_1(T \mid r_1)$ over the distribution of r_1 given by (3). We have

$$\begin{aligned} M_1(T) &= E(M_1(T \mid r_1)) \\ &= \frac{e^{(\alpha_1-\beta_1)(T-q_1)}\,(q_1-T)}{q_1-T-\nu_1 q_1+\nu_1 q_1 e^{(\alpha_1-\beta_1)(T-q_1)}} E(e^{r_1 \log \nu_1}) \\ &= \frac{e^{(\alpha_1-\beta_1)(T-q_1)}\,(q_1-T)}{q_1-T-q_1\nu_1+q_1\nu_1 e^{(\alpha_1-\beta_1)(T-q_1)}} \left[\frac{q_1\gamma}{\delta_2} e^{\log \nu_1}+\frac{\delta_1}{\delta_2}\{M_2(q_1(e^{\log \nu_1}-1))-M_2\}\right]. \end{aligned}$$

Hence, finally,

$$\begin{aligned} &\delta_2[q_1-T-q_1\nu_1+q_1\nu_1 e^{(\alpha_1-\beta_1)(T-q_1)}]\,M_1(T) \\ &\qquad = e^{(\alpha_1-\beta_1)(T-q_1)}\,(q_1-T)\,[q_1\gamma\nu_1+\delta_1\{M_2(q_1(\nu_1-1))-M_2\}]. \end{aligned} \tag{9}$$

Similarly,

$$\begin{aligned} &\delta_1[q_2-T-q_2\nu_2+q_2\nu_2 e^{(\alpha_2-\beta_2)(T-q_2)}]\,M_2(T) \\ &\qquad = e^{(\alpha_2-\beta_2)(T-q_2)}\,(q_2-T)\,[q_2\gamma\nu_2+\delta_2\{M_1(q_2(\nu_2-1))-M_1\}]. \end{aligned} \tag{10}$$

The solution of these equations in the unknown functions $M_1(T)$ and $M_2(T)$ to give the moments of the distributions of t_1 and t_2 has not been achieved in the general case. The solution has, however, been obtained when one of the β's is zero, and this is considered later. We note, however, that in the general case, expansion of (9) and (10) leads to

$$\left.\begin{aligned} &(1-\rho_1-\rho_2)\,\delta_2\mu_1'(t_1) = \delta_2\epsilon_1(1-\rho_2)+\delta_1\epsilon_2\rho_1+\rho_1\gamma \\ \text{and}\qquad &(1-\rho_1-\rho_2)\,\delta_1\mu_1'(t_2) = \delta_2\epsilon_1\rho_2+\delta_1\epsilon_2(1-\rho_1)+\rho_2\gamma. \end{aligned}\right\} \tag{11}$$

3·6. *Stage* 4

We now derive the mean delay to a V_1 by considering the average number of V_1's waiting at a random instant. Suppose that the average number waiting during a VB_1 is m_1 and during a VB_2 is m_2. Then the average number waiting at a random instant is

$$m = \frac{m_1\phi_1\mu_1'(t_1)+m_2\phi_2\mu_1'(t_2)}{\phi_0/(q_1+q_2)+\phi_1\mu_1'(t_1)+\phi_2\mu_1'(t_2)}, \tag{12}$$

since $\phi_0/(q_1+q_2)$, … are the relative amounts of time in VB_0's, …. The mean delay, which we denote by $\overline{w}_1$, is then

$$\overline{w}_1 = \frac{m}{q_1}. \tag{13}$$

In a VB_2 of length t_2, the average number of V_1's waiting will be $\frac{1}{2}q_1 t_2$, since the expected number at a time t after the beginning of the block is $q_1 t$. Hence, averaging over the distribution of t_2, we obtain

$$m_2 = \frac{E(\frac{1}{2}q_1 t_2 \,.\, t_2)}{E(t_2)} = \frac{\frac{1}{2}q_1\mu_2'(t_2)}{\mu_1'(t_2)}.$$

It thus remains to find m_1.

We find m_1 first of all in terms of $\psi(r)$, the expected total delay to V_1's during an r-wave, as previously defined. Consider a VB_1 which starts off with an r-wave. As before, we can split the block into a number of parts, during which the expected total delay can be written down immediately, as follows:

(i) During the r-wave, expected total delay $= \psi(r)$.

(ii) During the unit waves, expected total delay $= (e^{(\alpha_1-\beta_1)q_1}-1)\,\psi(1)$.

(iii) During the gaps, expected total delay $= 0$.

(iv) During the final $\alpha_1-\beta_1$, expected total delay $= 0$.

Hence, adding together for the whole block,

$$\text{expected delay total} = \psi(r) + (e^{(\alpha_1-\beta_1)q_1} - 1)\,\psi(1). \tag{14}$$

Averaging over the distribution of r, we obtain

$$\text{expected total delay} = \overline{\psi(r)} + (e^{(\alpha_1-\beta_1)q_1} - 1)\,\psi(1).$$

Also,

$$\text{average length of block} = \mu_1'(t_1).$$

m_1 is now obtained as the ratio of the last two expressions.

To find $\psi(r)$, we break up the r-wave into 'generations', as follows (the procedure is that followed by Borel, 1942): While the initial r vehicles are departing, suppose a further a arrive; and that while the a are departing, a further b arrive; and so on. The number of such generations can be indefinitely large, and the chance that the sizes of successive ones are $r, a, b, c, \ldots$ is evidently

$$\begin{aligned} P(a, b, \ldots \mid r) &= \frac{e^{-\rho_1 r}(\rho_1 r)^a}{a!}\,\frac{e^{-\rho_1 a}(\rho_1 a)^b}{b!}\,\frac{e^{-\rho_1 b}(\rho_1 b)^c}{c!}\ldots \\ &= \frac{e^{-\rho_1 r}\theta^{a+b+}\cdots r^a a^b b^c \ldots}{a!\,b!\,c!\ldots}, \quad \text{where} \quad \theta = \rho_1 e^{-\rho_1} \end{aligned} \tag{15}$$

(0^0 is to be taken as unity).

In this case, the expected total delay is

$$\tfrac{1}{2}\beta_1[r(r-1) + a(a-1) + b(b-1) + \ldots] + \tfrac{1}{2}\beta_1[ra + ab + bc + \ldots]. \tag{16}$$

The terms in the first bracket arise from delays to vehicles while vehicles in their own generation are departing, while those in the second arise from delays while the previous generation is departing.

The detailed algebra required to obtain the expectation of the terms in (16), averaged over the distribution given by (15), will be omitted, but may be obtained by noting, for example, that the probability of a particular a, b, c, d and e, followed by unspecified frequencies, is

$$e^{-\rho_1 r}\,\theta^{a+b+c+d}\rho_1^e r^a a^b b^c c^d d^e/(a!\,b!\,c!\,d!\,e!).$$

We find that, for example,

$$E(e(e-1)) = r(\rho_1^6 + \ldots + \rho_1^9) + r^2\rho_1^{10}$$

and

$$E(de) \quad = r(\rho_1^5 + \ldots + \rho_1^8) + r^2\rho_1^9.$$

We finally obtain the expected total delay as

$$\psi(r) = \tfrac{1}{2}\beta_1\left[\frac{r(2\rho_1 - 1)}{(1-\rho_1)^2} + \frac{r^2}{1-\rho_1}\right].$$

Thus

$$\psi(1) = \tfrac{1}{2}\beta_1\frac{\rho_1}{(1-\rho_1)^2},$$

$$\overline{\psi(r)} = \tfrac{1}{2}\beta_1\left[\frac{2\rho_1 - 1}{(1-\rho_1)^2}\mu_1'(r) + \frac{1}{1-\rho_1}\,\mu_2'(r)\right].$$

Alternatively, it may be shown quite simply that the joint m.g.f. of $a, b, c, \ldots$, defined as

$$E(e^{at_1 + bt_2 + ct_3 + \ldots}),$$

is equal to

$$\exp\,(-\rho_1 r + r\theta \exp\,(t_1 + \theta \exp\,(t_2 + \theta \exp\,(\ldots)))).$$

(This is closely related to expressions given by Good, 1948.) Expansion of this function as a power series in $t_1, t_2, \ldots$ then leads to the same value of $\psi(r)$.

Hence, averaged over the distribution of r, the expected total delay to V_1's during a VB_1 is

$$\tfrac{1}{2}\beta_1\left[\frac{2\rho_1-1}{(1-\rho_1)^2}\mu_1'(r)+\frac{1}{1-\rho_1}\mu_2'(r)\right]+[e^{(\alpha_1-\beta_1)q_1}-1]\,\tfrac{1}{2}\beta_1\frac{\rho_1}{(1-\rho_1)^2}. \tag{17}$$

Expansion of the generating function (3) gives

$$\mu_1'(r) = q_1\gamma/\delta_2+\delta_1 q_1\mu_1'(t_2)/\delta_2$$

and

$$\mu_2'(r) = q_1\gamma/\delta_2+\delta_1[q_1\mu_1'(t_2)+q_1^2\mu_2'(t_2)]/\delta_2.$$

Substituting these expressions in (17) and the expression in (11) for $\mu_1'(t_2)$ then gives the expected total delay as

$$\frac{\tfrac{1}{2}\rho_1 q_1}{(1-\rho_1)(1-\rho_1-\rho_2)\,\delta_2}\,[\beta_1(\gamma+\epsilon_1\delta_2+\epsilon_2\delta_1)+\delta_1(1-\rho_1-\rho_2)\,\mu_2'(t_2)].$$

Collecting our results together, and substituting in (12) and (13), we find that the mean delay to V_1's is

$$\overline{w}_1 = \frac{1}{2(1-\rho_1)}\left[\beta_1\rho_1+\frac{\delta_1\mu_2'(t_2)\,(1-\rho_1-\rho_2)}{\gamma+\epsilon_1\delta_2+\epsilon_2\delta_1}\right]. \tag{18}$$

Similarly

$$\overline{w}_2 = \frac{1}{2(1-\rho_2)}\left[\beta_2\rho_2+\frac{\delta_2\mu_2'(t_1)\,(1-\rho_1-\rho_2)}{\gamma+\epsilon_1\delta_2+\epsilon_2\delta_1}\right]. \tag{19}$$

This completes stage 4.

4. EXPLICIT SOLUTIONS FOR SOME SPECIAL CASES

As already mentioned, the analysis of § 3 leads to equations for the mean delays whose solutions have not been obtained. If, however, we assume one of the β's to be zero, an explicit solution can be found, and we now consider this case.

4·1. *Special case $\beta_2 = 0$*

In this case, equation (10) reduces to

$$M_2(T) = \frac{q_2-T}{q_2-Te^{\alpha_2(q_2-T)}}, \tag{20}$$

so that

$$M_2 = M_2(-q_1) = \frac{q_1+q_2}{q_2+q_1e^{\alpha_2(q_1+q_2)}}. \tag{21}$$

Substituting in (9), and putting $T=-q_2$, leads on simplification to

$$M_1 = \frac{q_1+q_2}{q_1+e^{(\alpha_1-\beta_1)(q_1+q_2)}\dfrac{q_2-q_1(\omega-1)\,e^{\alpha_2(q_1+q_2-q_1\omega)}}{1+(\omega-1)\,e^{-\alpha_2q_1\omega}}}, \tag{22}$$

where ω is the smaller real root of

$$\omega = e^{-\beta_1(q_1+q_2-q_1\omega)}.$$

Expansion of equations (9) and (20) leads to

$$\left.\begin{aligned}\delta_2q_1q_2(1-\rho_1)^3\,\mu_2'(t_1) &= 2\delta_1\rho_1^2(1-\rho_1)\,q_1(1+\epsilon_2q_2)\,(\epsilon_2-\alpha_2)\\ &\quad+2\delta_2q_2(1-\rho_1)^2\,(1+\epsilon_1q_1)\,(\epsilon_1-\alpha_1+\beta_1)\\ &\quad+q_2\rho_1(\gamma+\delta_1\epsilon_2+\delta_2\epsilon_1)\,(\rho_1+2q_1\epsilon_1-2q_1\epsilon_1\rho_1) = P,\\ q_2\mu_2'(t_2) &= 2(1+\epsilon_2q_2)\,(\epsilon_2-\alpha_2)\end{aligned}\right\} \tag{23}$$

Hence, by (18) and (19),

$$\begin{aligned}\overline{w}_1 &= \frac{\rho_1^2}{2q_1(1-\rho_1)} + \frac{\delta_1(\epsilon_2 q_2 + 1)(\epsilon_2 - \alpha_2)}{q_2(\gamma + \epsilon_1\delta_2 + \epsilon_2\delta_1)} \\ &= \frac{\rho_1^2}{2q_1(1-\rho_1)} + \frac{\delta_1 e^{\alpha_2 q_2}(e^{\alpha_2 q_2} - \alpha_2 q_2 - 1)}{q_2^2(\gamma + \epsilon_1\delta_2 + \epsilon_2\delta_1)},\end{aligned} \tag{24}$$

and

$$\overline{w}_2 = \frac{1}{2(1-\rho_1)^2 q_1 q_2} \frac{P}{\gamma + \epsilon_1\delta_2 + \epsilon_2\delta_1}, \quad \text{where } P \text{ is given by (23).} \tag{25}$$

These expressions for $\overline{w}_1$ and $\overline{w}_2$ are explicit formulae, since the only quantities they contain, apart from the basic constants, are γ, δ_1 and δ_2, which can be expressed in terms of these by means of the relation (21) and (22). However, they are obviously unsuitable for tabulation and it is not thought that any important simplification can be made to them. This special case will therefore be left in this rather unsatisfactory state.

4·2. *Special case $\beta_1 = \beta_2 = 0$*

From (24) and (25) we have

$$\overline{w}_1 = \frac{\delta_1 e^{\alpha_2 q_2}(e^{\alpha_2 q_2} - \alpha_2 q_2 - 1)}{q_2^2(\gamma + \epsilon_1\delta_2 + \epsilon_2\delta_1)},$$

$$\overline{w}_2 = \frac{\delta_2 e^{\alpha_1 q_1}(e^{\alpha_1 q_1} - \alpha_1 q_1 - 1)}{q_1^2(\gamma + \epsilon_1\delta_2 + \epsilon_2\delta_1)}.$$

Using (21) and (22) we find

$$\gamma = \frac{(q_1+q_2)(q_2 e^{\alpha_1(q_1+q_2)} + q_1 e^{\alpha_2(q_1+q_2)})}{(q_2 + q_1 e^{\alpha_2(q_1+q_2)})(q_1 + q_2 e^{\alpha_1(q_1+q_2)})},$$

$$\delta_1 = \frac{(q_1+q_2)\, q_2 e^{\alpha_1(q_1+q_2)}}{q_1 + q_2 e^{\alpha_1(q_1+q_2)}}, \quad \delta_2 = \frac{(q_1+q_2)\, q_1 e^{\alpha_2(q_1+q_2)}}{q_2 + q_1 e^{\alpha_2(q_1+q_2)}},$$

whence

$$\overline{w}_1 = \frac{e^{\alpha_1 q_2}(q_2 + q_1 e^{\alpha_2(q_1+q_2)})}{q_2 e^{\alpha_1 q_2}(e^{(\alpha_1+\alpha_2)q_1} - e^{\alpha_2 q_1} + 1) + q_1 e^{\alpha_2 q_1}(e^{(\alpha_1+\alpha_2)q_2} - e^{\alpha_1 q_2} + 1)} \frac{e^{\alpha_2 q_2} - \alpha_2 q_2 - 1}{q_2} \tag{26}$$

and

$$\overline{w}_2 = \frac{e^{\alpha_2 q_1}(q_1 + q_2 e^{\alpha_1(q_1+q_2)})}{q_1 e^{\alpha_2 q_1}(e^{(\alpha_1+\alpha_2)q_2} - e^{\alpha_1 q_2} + 1) + q_2 e^{\alpha_1 q_2}(e^{(\alpha_1+\alpha_2)q_1} - e^{\alpha_2 q_1} + 1)} \frac{e^{\alpha_1 q_1} - \alpha_1 q_1 - 1}{q_1}. \tag{27}$$

Table 1 below tabulates $\overline{w}_1$ for the case $\alpha_1 = \alpha_2 = 1$, for $q_1 = 0{\cdot}0\,(0{\cdot}5)\,2{\cdot}0$, $q_2 = 0{\cdot}0\,(0{\cdot}5)\,2{\cdot}0$.

Table 1. *Values of $\overline{w}_1$ for $\beta_1 = \beta_2 = 0$, $\alpha_1 = \alpha_2 = 1$*

		q_1				
		0·0	0·5	1·0	1·5	2·0
q_2	0·0	0·00	0·00	0·00	0·00	0·00
	0·5	0·30	0·27	0·24	0·21	0·17
	1·0	0·72	0·61	0·53	0·45	0·37
	1·5	1·32	1·11	0·97	0·84	0·70
	2·0	2·20	1·87	1·67	1·47	1·27

4·3. *Special case $\alpha_2 = \beta_2 = 0$*

In this case we have $M_2 = 1, \quad \delta_2 = q_1, \quad \epsilon_2 = 0, \quad \gamma = 1.$
Hence by (24) and (25),

$$\overline{w}_1 = \frac{\rho_1^2}{2q_1(1-\rho_1)}, \tag{28}$$

$$\begin{aligned}\overline{w}_2 &= \frac{1}{2(1-\rho_1)^2 q_1 q_2} q_2[2\rho_1 - 3\rho_1^2 + 2\rho_1^3 + 2q_1\epsilon_1(1-\rho_1) - 2q_1\alpha_1(1-\rho_1)^2] \\ &= \frac{e^{(\alpha_1-\beta_1)q_1}}{q_1(1-\rho_1)} - \alpha_1 - \frac{1}{q_1} + \frac{-\rho_1^2 + 2\rho_1^3}{2q_1(1-\rho_1)^2}.\end{aligned} \tag{29}$$

This case corresponds to a physical situation in which vehicles in stream 2 do not affect those in stream 1 in any way. $\overline{w}_1$ is therefore the mean delay for a simple queue with arrivals q_1 per unit time and service time β_1.

Table 2 below tabulates $\overline{w}_2$ for $\alpha_1 = 1$, $\beta_1 = 0{\cdot}0\,(0{\cdot}1)\,1{\cdot}0$, $q_1 = 0{\cdot}0\,(0{\cdot}2)\,2{\cdot}0$, for values of $\overline{w}_2$ less than 10.

Table 2. *Values of $\overline{w}_2$ for the special case $\alpha_2 = \beta_2 = 0$, $\alpha_1 = 1$*

		β_1										
		0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1·0
q_1	0·0	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00
	0·2	0·11	0·11	0·11	0·11	0·11	0·12	0·12	0·13	0·13	0·14	0·16
	0·4	0·23	0·23	0·23	0·24	0·25	0·27	0·30	0·34	0·38	0·46	0·56
	0·6	0·37	0·37	0·38	0·40	0·43	0·49	0·57	0·70	0·92	1·27	1·88
	0·8	0·53	0·54	0·56	0·60	0·67	0·83	1·04	1·37	2·37	4·40	10·00
	1·0	0·72	0·73	0·76	0·84	0·99	1·30	1·96	3·59	8·91	—	*
	1·2	0·93	0·95	1·01	1·15	1·44	2·15	4·19	—	—	*	*
	1·4	1·18	1·21	1·30	1·54	2·12	3·86	—	—	*	*	*
	1·6	1·47	1·51	1·66	2·05	3·19	8·33	—	*	*	*	*
	1·8	1·80	1·86	2·08	2·73	5·10	—	*	*	*	*	*
	2·0	2·20	2·27	2·60	3·56	9·20	*	*	*	*	*	*

* $\rho_1 \geqslant 1$; no stationary solution exists.

In this special case, it is possible to obtain an expression for the m.g.f. $M_{w_2}(T)$ of the distribution of the delay w_2. We proceed as follows:

If a V_2 arrives in a VB_1 of length t_1, its delay will have a rectangular distribution in the interval $(0, t_1)$. The m.g.f. of this delay will therefore be

$$(e^{t_1 T} - 1)/t_1 T.$$

Averaging over the distribution of t_1 gives

$$E[t_1(e^{t_1 T} - 1)/t_1 T]/E[t_1] = [M_1(T) - 1]/T\mu_1'(t_1). \tag{30}$$

Now, by (7), and since r_1 is always unity, we have

$$M_1(T) = \frac{\nu e^{(\alpha_1-\beta_1)(T-q_1)}(q_1 - T)}{q_1 - T - \nu_1 q_1 + \nu_1 q_1 e^{(\alpha_1-\beta_1)(T-q_1)}},$$

so that

$$\mu_1'(t_1) = \frac{e^{(\alpha_1-\beta_1)q_1}}{q_1(1-\rho_1)} - \frac{1}{q_1}. \tag{31}$$

If a V_2 arrives in a VB_0, it will have zero delay. The probability of this is

$$\frac{1/q_1}{\mu_1'(t_1)+1/q_1} = \frac{1-\rho_1}{e^{(\alpha_1-\beta_1)q_1}}, \tag{32}$$

since VB_1's and VB_0's are equally frequent.

From (30), (31) and (32) we obtain

$$M_{w_2}(T) = \frac{(1-\rho_1)(q_1-T)(\nu_1 q_1 - q_1 + T)}{Te^{(\alpha_1-\beta_1)q_1}[-\nu_1 q_1 + q_1 - T + \nu_1 q_1 e^{(\alpha_1-\beta_1)(T-q_1)}]}. \tag{33}$$

4·4. *The special case $\beta_1 = \alpha_2 = \beta_2 = 0$*

The solution for this case can be derived quite easily from the last section. We obtain

$$\left.\begin{aligned} M_{w_2}(T) &= \frac{(q_1-T)\,e^{-\alpha_1 T-\alpha_1 q_1}}{q_1 e^{-\alpha_1 q_1} - Te^{-\alpha_1 T}},\\ \text{and}\qquad \overline{w}_1 &= 0,\\ \overline{w}_2 &= (e^{\alpha_1 q_1} - \alpha_1 q_1 - 1)/q_1. \end{aligned}\right\} \tag{34}$$

This case has been considered in more detail by Garwood (1940), Raff (1951) and Tanner (1951).

5. APPLICATIONS

In this section we consider briefly some particular applications of the preceding analysis.

5·1. *Delays to pedestrians crossing a road*

Suppose a stream of vehicles, flow q_1 per unit time, is passing a bottleneck where they are constrained to pass in a single line, with minimum separation β_1. At this bottleneck, pedestrians, who demand a gap α_1 between themselves and the next vehicle to arrive, are crossing the road. Then their average delay is given by (29) and Table 2. This can be seen by noting that both for the situation just stated and for the solution of the general problem with $\alpha_1 > \beta_1 > 0$, $\alpha_2 = \beta_2 = 0$, vehicles in stream 1 are unaffected by those in stream 2, while vehicles (or pedestrians) in stream 2 wait for a gap α_1 in stream 1 and do not hold it up. (It is assumed that the time taken by a pedestrian to cross the road is less than α_1, so that vehicles are not held up by pedestrians.) When there is no bottleneck, i.e. $\beta_1 = 0$, the delay is given by (34).

5·2. *Delays to vehicles on minor road at uncontrolled intersections*

If the suffix 2 refers to vehicles using the minor road, and we suppose that these vehicles have to cross or merge with just one main road stream, vehicles in the latter having absolute right of way, we have to consider the situation $\alpha_1 > \beta_1 > 0$, $\alpha_2 = 0$, $\beta_2 > 0$. (The same type of reasoning holds as in the previous paragraph.) This is a specialization of case (iii), as defined earlier, and the solution has not been obtained. If, however, the minor road traffic flow is small, it will not affect the result much if we put $\beta_2 = 0$, in which case the mean delay is given by (29), as for the pedestrian case above.

5·3. *Delays to vehicles at uncontrolled intersections where neither stream has absolute right of way*

If we consider just two streams of traffic intersecting at right angles, the mathematical problem is that of case (i), $\alpha_1 > \beta_1 > 0$, $\alpha_2 > \beta_2 > 0$. The solution for this has not been found, but, as in the previous section, we can give the mean delays if β_1 and β_2 (or just one of them) are negligible, or if one or both of the flows is very light. The delays are then given by (24) and (25), or by (26) and (27).

5·4. *Delays to intersecting streams of pedestrians*

There are a number of similar problems under this heading.

If we have a staircase, passage, doorway, gangway, bridge, etc., on which one-way traffic is enforced, we have a problem mathematically identical with the vehicle delay problem in terms of which this paper has been formulated. The simplification $\beta_1 = \beta_2 = 0$ may more often be appropriate here than in vehicle problems.

There will in practice often be complications in these situations, such as multi-channel operation, non-random arrivals, and so forth. An interesting, though no doubt very difficult problem, which has its counterpart in the vehicle delay problems, is to find what happens if instead of having single-channel operation we have several channels which could be operated some in one direction, some in the other, at the same time. Passing from one channel to another in the middle of the passage through it might or might not be permitted.

This paper was prepared as part of the programme of the Road Research Board of the Department of Scientific and Industrial Research, and is published by permission of the Director of Road Research.

The author wishes to thank Dr F. Garwood and Mr D. G. Kendall for advice in its preparation.

REFERENCES

BOREL, E. (1942). *C.R. Acad. Sci., Paris*, **214**, 452–6.
GARWOOD, F. (1940). *Suppl. J. R. Statist. Soc.* **7**, 65–77.
GOOD, I. J. (1948). *Proc. Camb. Phil. Soc.* **45**, 360–3.
HARRIS, T. E. (1948). *Ann. Math. Statist.* **19**, 474–94.
KENDALL, D. G. (1951). *J. R. Statist. Soc.* B, **13**, 151–85.
RAFF, M. S. (1951). *J. Amer. Statist. Assoc.* **46**, 253.
TANNER, J. C. (1951). *Biometrika*, **38**, 383–92.

Table 2. *Angular transformation of percentages less than two*

100p	0·00	0·01	0·02	0·03	0·04	0·05	0·06	0·07	0·08	0·09	1	2	3	4	5
0·0	0·327	0·960	1·222	1·423	1·592	1·741	1·876	2·001	2·116	2·225	} See note below				
0·1	2·327	2·425	2·518	2·608	2·694	2·777	2·858	2·935	3·011	3·085					
0·2	3·156	3·226	3·295	3·361	3·427	3·491	3·553	3·615	3·675	3·735	6	13	19	26	32
0·3	3·793	3·850	3·907	3·962	4·017	4·071	4·124	4·177	4·228	4·279	5	11	16	22	27
0·4	4·330	4·380	4·429	4·477	4·525	4·573	4·620	4·666	4·712	4·758	5	10	14	19	24
0·5	4·803	4·848	4·892	4·935	4·979	5·022	5·064	5·107	5·148	5·190	4	9	13	17	22
0·6	5·231	5·272	5·312	5·352	5·392	5·432	5·471	5·510	5·548	5·587	4	8	12	16	20
0·7	5·625	5·663	5·700	5·737	5·774	5·811	5·848	5·884	5·920	5·956	4	7	11	15	18
0·8	5·991	6·027	6·062	6·097	6·132	6·166	6·201	6·235	6·269	6·303	3	7	10	14	17
0·9	6·336	6·369	6·403	6·436	6·469	6·501	6·534	6·566	6·598	6·630	3	7	10	13	16
1·0	6·662	6·694	6·725	6·757	6·788	6·819	6·850	6·881	6·912	6·942	3	6	9	12	16
1·1	6·973	7·003	7·033	7·063	7·093	7·122	7·152	7·182	7·211	7·240	3	6	9	12	15
1·2	7·269	7·298	7·327	7·356	7·384	7·413	7·441	7·470	7·498	7·526	3	6	9	11	14
1·3	7·554	7·582	7·609	7·637	7·665	7·692	7·720	7·747	7·774	7·801	3	5	8	11	14
1·4	7·828	7·855	7·882	7·908	7·935	7·961	7·988	8·014	8·040	8·066	3	5	8	11	13
1·5	8·093	8·118	8·144	8·170	8·196	8·222	8·247	8·273	8·298	8·323	3	5	8	10	13
1·6	8·349	8·374	8·399	8·424	8·449	8·474	8·498	8·523	8·548	8·572	2	5	7	10	12
1·7	8·597	8·621	8·646	8·670	8·694	8·718	8·742	8·766	8·790	8·814	2	5	7	10	12
1·8	8·838	8·862	8·885	8·909	8·933	8·956	8·980	9·003	9·026	9·050	2	5	7	9	12
1·9	9·073	9·096	9·119	9·142	9·165	9·188	9·211	9·233	9·256	9·279	2	5	7	9	11

Interpolate linearly between tabular values when $0{\cdot}05 < 100p < 0{\cdot}2$. When $100p < 0{\cdot}05$, use the formula, $\theta = 0{\cdot}327 + 6{\cdot}325\sqrt{p}$. For values of $100p$ greater than 98, enter with $100(1-p)$ and subtract the tabular value from 100.

Table 3. *Angular transformation of proper fractions*

*1/30	11·939	*1/10	20·676	5/27	28·462	8/29	35·301	*11/30	41·464	9/20	46·832
1/29	12·140	3/29	21·037	3/16	28·650	5/18	35·436	7/19	41·579	5/11	47·121
1/28	12·352	2/19	21·225	4/21	28·891	7/25	35·593	10/27	41·707	11/24	47·362
1/27	12·575	3/28	21·418	5/26	29·038	2/7	35·994	3/8	42·010	6/13	47·565
1/26	12·811	1/9	21·820	*1/5	29·651	7/24	36·409	11/29	42·291	13/28	47·739
1/25	13·062	3/26	22·247	6/29	30·192	5/17	36·580	8/21	42·398	*7/15	47·890
1/24	13·328	2/17	22·470	5/24	30·304	8/27	36·731	5/13	42·636	8/17	48·139
1/23	13·612	3/25	22·700	4/19	30·475	*3/10	36·987	7/18	42·914	9/19	48·335
1/22	13·915	1/8	23·182	3/14	30·766	7/23	37·286	9/23	43·070	10/21	48·494
1/21	14·240	3/23	23·697	5/23	31·004	4/13	37·516	11/28	43·171	11/23	48·625
1/20	14·590	*2/15	23·968	2/9	31·373	9/29	37·697	*2/5	43·633	12/25	48·735
1/19	14·967	3/22	24·248	5/22	31·756	5/16	37·844	11/27	44·110	13/27	48·829
1/18	15·376	4/29	24·392	3/13	32·019	6/19	38·068	9/22	44·218	14/29	48·909
1/17	15·821	1/7	24·841	*7/30	32·211	7/22	38·231	7/17	44·390	*1/2	50·000
1/16	16·308	4/27	25·315	4/17	32·357	8/25	38·354	12/29	44·520		
*1/15	16·844	3/20	25·480	5/21	32·566	9/28	38·451	5/12	44·705		
2/29	17·133	2/13	25·819	6/25	32·707	*1/3	39·253	8/19	44·986		
1/14	17·438	3/19	26·172	7/29	32·809	10/29	40·021	11/26	45·116		
2/27	17·760	4/25	26·354	1/4	33·442	9/26	40·109	3/7	45·467		
1/13	18·101	*1/6	26·924	7/27	34·114	8/23	40·221	*13/30	45·771		
2/25	18·463	5/29	27·408	6/23	34·231	7/20	40·365	10/23	45·864		
1/12	18·848	4/23	27·534	5/19	34·395	6/17	40·560	7/16	46·037		
2/23	19·259	3/17	27·746	*4/15	34·647	5/14	40·837	11/25	46·196		
1/11	19·698	5/28	27·920	7/26	34·830	9/25	41·026	4/9	46·479		
2/21	20·169	2/11	28·187	3/11	35·078	4/11	41·265	13/29	46·723		

For fractions exceeding one-half, subtract the fraction from unity and the angle from one hundred. Each thirtieth is marked by an asterisk.

None of these advantages can be considered as overwhelming, but, since a new tabulation was to be undertaken, they were sufficient to lead to the choice of this form for the transformation. As an alternative, we also considered taking θ as one-tenth of the expression above but the resultant table would look so much like a table of probits that it would not be at all difficult to enter one in mistake for the other.

Table 1 gives θ for $p = 0\,(0{\cdot}1)\,50$ with proportional parts for linear interpolation. If $100p > 50$, enter the table with $100(1-p)$ and subtract the tabular value from 100. For example, if $100p = 62{\cdot}77$, we enter with 37·23 and finding that $\theta = 41{\cdot}833$ obtain for the required transformed variable a value of 58·167. Table 2 gives a finer tabulation for the first 2 %, $p = 0\,(0{\cdot}01)\,2$ %, and hence for the last 2 % of the range, although we note that it will not usually be profitable to use the angular transformation when p is close to 0 or 100. Table 3, following the pattern of the original table in Fisher & Yates, gives the angles corresponding to proper fractions with denominators not exceeding 30.

The writer has also calculated a table of angular values for adjustments of special accuracy, analogous to Table XIV given by Fisher & Yates (1938). This table is not reproduced here.

The tables were computed with the aid of *British Association Mathematical Tables*, vol. 1 (1931).

REFERENCES

BARTLETT, M. S. (1937). Subsampling for attributes. *Suppl. J. R. Statist. Soc.* **4**, 131–5.

BLISS, C. I. (1937). The analysis of field experimental data expressed in percentages. *Pl. Prot., Leningr.*, **12**, 70–72 (in Russian).

British Association Mathematical Tables (1931). I. Circular and Hyperbolic Functions. Cambridge University Press.

EISENHART, C., HASTAY, M. W. & WALLIS, W. A. (1947). *Selected Techniques of Statistical Analysis.* New York: McGraw-Hill Book Co. Inc.

FISHER, R. A. (1922). On the dominance ratio. *Proc. Roy. Soc. Edinb.* **42**, 321–41.

FISHER, R. A. & YATES, F. (1938). *Statistical Tables for Biological, Agricultural and Medical Research.* London: Oliver and Boyd.

SNEDECOR, G. W. (1946). *Statistical Methods.* The Iowa State College Press.

TESTS OF SIGNIFICANCE IN A 2×2 CONTINGENCY TABLE: EXTENSION OF FINNEY'S TABLE

COMPUTED BY R. LATSCHA

Institute of Actuarial Science, University of Berne

EDITORIAL FOREWORD

Finney (1948) has given a table which may be used to test the significance of the deviation from proportionality in any 2 × 2 contingency table having both the frequencies in one of its margins less than or equal to 15. The table printed below extends the range of Finney's table up to marginal frequencies of 20. As the interpretation and uses of the new table are exactly similar to those of the 1948 table, only a brief introductory statement is required.*

Using Finney's notation, the contingency table should be arranged in the form

	Number of		Total
	Successes	Failures	
Series I	a	$A-a$	A
Series II	b	$B-b$	B
	$r=a+b$	$A+B-a-b$	$A+B$

where series I is defined to be that which makes $A \geqslant B$, and the type of observation conventionally regarded as a 'success' is that which makes $a/A \geqslant b/B$. The table of significance levels is arranged in sections according to the value of A; the sections for $A = 9(1)15$ were given by Finney, while those for $A = 16(1)20$ computed by Latscha are printed below.

For given data, the table is entered in the section for A, the subsection for B and the line for a; then the main body of the table shows in bold type the appropriate significance points for b. Thus if the observed value of b is equal to or less than the bold integer in the column headed 0·05, 0·025, 0·01 or 0·005, then a/A is significantly greater than b/B (single-tail test) at these probability levels. On the other hand, for the two-tail test, if b is equal to or less than the integer in a given column, a/A is significantly *different from* b/B at a probability level equal to *twice* the figure heading that column, i.e. at the 0·10, 0·05, 0·02 and 0·01 levels, respectively. A dash, or absence of an entry, for some combination of A, B and a indicates that no 2 × 2 table in that class can show a significant effect at that level.

Owing to the discontinuous character of the hypergeometric distribution, the conditional probability that, for a given value of $a+b$, b will be equal to or less than the value specified in bold type will generally be less, and often very considerably less, than that shown at the head of the column; the true probabilities are given in small type.

---

* Copies of Finney's table are available as a *Biometrika* 'separate' and the present extension will be made available in similar form. Finney's table, but not the extension, is included in the new *Biometrika Tables for Statisticians* now at Press.

As an illustration, we may use Lange's data on criminality among twin brothers or sisters of criminals (Fisher, 1946, §21·01). This example was taken by Finney (1948), but as $A > 15$ he used it to show how his table could be extended under certain conditions. As $A \leqslant 20$, direct entry is now possible in Latscha's table.

The contingency table shows the number of twin brother or sisters of criminals who had also been convicted of crime, classing separately monozygotic and dizygotic (but like-sexed) twins:

	Not convicted	Convicted	Total
Dizygotic	15 (= a)	2	17 (= A)
Monozygotic	3 (= b)	10	13 (= B)
Total	18	12	30

Following the rule given above, the letters A, B, a and b are associated with the observed frequencies as shown. The null-hypothesis is that the twin of a criminal is no more likely to be convicted of crime if the twinning is monozygotic than if it is dizygotic. If the only deviation from the hypothesis which we are prepared to consider is that monozygotic twins behave more similarly than dizygotic, a single-tail test will be appropriate and we shall ask whether $a/A = 15/17$ is significantly greater than $b/B = 3/13$.

Turning to the appropriate section of the table with $A = 17$, $B = 13$ and $a = 15$ we find that the observed value of $b = 3$ is significant at the 0·5 % level, since it is less than 4, the last entry in the row of bold figures.

The figure in small type following $b = 4$ indicates that for a contingency table with marginal frequencies

$$A = 17, \quad B = 13, \quad r = a+b = 15+4 = 19, \quad A+B-a-b = 11,$$

the conditional probability of an arrangement within the table with $b \leqslant 4$ is 0·002 on the null hypothesis of independence. The probability that $b \leqslant 3$ within the observed table (having $a+b = 18$) is not recorded, but is $< 0{\cdot}002$.

As far as possible, checks on the internal consistency of the table have been made as well as comparisons with the more extensive tables for the special case $A = B$ published by Swaroop (1950).

REFERENCES

FINNEY, D. J. (1948). *Biometrika*, **35**, 148.
FISHER, R. A. (1946). *Statistical Methods for Research Workers*, 10th ed. Edinburgh: Oliver and Boyd.
SWAROOP, S. (1950). *Indian Med. Res. Mem.*, no. 35.

	a	Probability			
		0·05	0·025	0·01	0·005
A=16 B=16	16	**11** ·022	**11** ·022	**10** ·009	**9** ·003
	15	**10** ·041	**9** ·019	**8** ·008	**7** ·003
	14	**8** ·027	**7** ·012	**6** ·005^{-}	**6** ·005^{-}
	13	**7** ·033	**6** ·015^{-}	**5** ·006	**4** ·002
	12	**6** ·037	**5** ·016	**4** ·006	**3** ·002
	11	**5** ·038	**4** ·016	**3** ·006	**2** ·002
	10	**4** ·037	**3** ·015^{-}	**2** ·005^{-}	**2** ·005^{-}
	9	**3** ·033	**2** ·012	**1** ·003	**1** ·003
	8	**2** ·027	**1** ·008	**1** ·008	**0** ·001
	7	**1** ·019	**1** ·019	**0** ·003	**0** ·003
	6	**1** ·041	**0** ·009	**0** ·009	—
	5	**0** ·022	**0** ·022	—	—
15	16	**11** ·043	**10** ·018	**9** ·007	**8** ·002
	15	**9** ·033	**8** ·014	**7** ·005^{+}	**6** ·002
	14	**8** ·044	**7** ·019	**6** ·008	**5** ·003
	13	**6** ·023	**6** ·023	**5** ·009	**4** ·003
	12	**5** ·024	**5** ·024	**4** ·009	**3** ·003
	11	**4** ·023	**4** ·023	**3** ·008	**2** ·002
	10	**4** ·049	**3** ·020	**2** ·006	**1** ·001
	9	**3** ·043	**2** ·016	**1** ·004	**1** ·004
	8	**2** ·035^{-}	**1** ·010^{+}	**0** ·002	**0** ·002
	7	**1** ·023	**1** ·023	**0** ·004	**0** ·004
	6	**0** ·011	**0** ·011	—	—
	5	**0** ·026	—	—	—
14	16	**10** ·037	**9** ·014	**8** ·005^{+}	**7** ·002
	15	**8** ·025^{+}	**7** ·010^{-}	**7** ·010^{-}	**6** ·003
	14	**7** ·032	**6** ·013	**5** ·005^{-}	**5** ·005^{-}
	13	**6** ·035^{+}	**5** ·014	**4** ·005^{+}	**3** ·001
	12	**5** ·035^{+}	**4** ·014	**3** ·005^{-}	**3** ·005^{-}
	11	**4** ·033	**3** ·012	**2** ·004	**2** ·004
	10	**3** ·028	**2** ·009	**2** ·009	**1** ·002
	9	**2** ·021	**2** ·021	**1** ·006	**0** ·001
	8	**2** ·045^{-}	**1** ·013	**0** ·002	**0** ·002
	7	**1** ·030	**0** ·006	**0** ·006	—
	6	**0** ·013	**0** ·013	—	—
	5	**0** ·031	—	—	—
13	16	**9** ·030	**8** ·011	**7** ·004	**7** ·004
	15	**8** ·047	**7** ·019	**6** ·007	**5** ·002
	14	**6** ·023	**6** ·023	**5** ·008	**4** ·003
	13	**5** ·023	**5** ·023	**4** ·008	**3** ·003
	12	**4** ·022	**4** ·022	**3** ·007	**2** ·002
	11	**4** ·048	**3** ·018	**2** ·005^{+}	**1** ·001
	10	**3** ·039	**2** ·013	**1** ·003	**1** ·003
	9	**2** ·029	**1** ·008	**1** ·008	**0** ·001
	8	**1** ·018	**1** ·018	**0** ·003	**0** ·003
	7	**1** ·038	**0** ·007	**0** ·007	—
	6	**0** ·017	**0** ·017	—	—
	5	**0** ·037	—	—	—

	a	Probability			
		0·05	0·025	0·01	0·005
A=16 B=12	16	**8** ·024	**8** ·024	**7** ·008	**6** ·002
	15	**7** ·036	**6** ·013	**5** ·004	**5** ·004
	14	**6** ·040	**5** ·015^{-}	**4** ·005^{-}	**4** ·005^{-}
	13	**5** ·039	**4** ·014	**3** ·004	**3** ·004
	12	**4** ·034	**3** ·012	**2** ·003	**2** ·003
	11	**3** ·027	**2** ·008	**2** ·008	**1** ·002
	10	**2** ·019	**2** ·019	**1** ·005^{-}	**1** ·005^{-}
	9	**2** ·040	**1** ·011	**0** ·002	**0** ·002
	8	**1** ·024	**1** ·024	**0** ·004	**0** ·004
	7	**1** ·048	**0** ·010^{-}	**0** ·010^{-}	—
	6	**0** ·021	**0** ·021	—	—
	5	**0** ·044	—	—	—
11	16	**7** ·019	**7** ·019	**6** ·006	**5** ·002
	15	**6** ·027	**5** ·009	**5** ·009	**4** ·002
	14	**5** ·027	**4** ·009	**4** ·009	**3** ·002
	13	**4** ·024	**4** ·024	**3** ·008	**2** ·002
	12	**3** ·019	**3** ·019	**2** ·005^{+}	**1** ·001
	11	**3** ·041	**2** ·013	**1** ·003	**1** ·003
	10	**2** ·028	**1** ·007	**1** ·007	**0** ·001
	9	**1** ·016	**1** ·016	**0** ·002	**0** ·002
	8	**1** ·033	**0** ·006	**0** ·006	—
	7	**0** ·013	**0** ·013	—	—
	6	**0** ·027	—	—	—
10	16	**7** ·046	**6** ·014	**5** ·004	**5** ·004
	15	**5** ·018	**5** ·018	**4** ·005^{+}	**3** ·001
	14	**4** ·018	**4** ·018	**3** ·005^{-}	**3** ·005^{-}
	13	**4** ·042	**3** ·014	**2** ·003	**2** ·003
	12	**3** ·032	**2** ·009	**2** ·009	**1** ·002
	11	**2** ·021	**2** ·021	**1** ·005^{-}	**1** ·005^{-}
	10	**2** ·042	**1** ·011	**0** ·002	**0** ·002
	9	**1** ·023	**1** ·023	**0** ·004	**0** ·004
	8	**1** ·045^{-}	**0** ·008	**0** ·008	—
	7	**0** ·017	**0** ·017	—	—
	6	**0** ·035^{-}	—	—	—
9	16	**6** ·037	**5** ·010^{-}	**5** ·010^{-}	**4** ·002
	15	**5** ·040	**4** ·012	**3** ·003	**3** ·003
	14	**4** ·034	**3** ·010^{-}	**3** ·010^{-}	**2** ·002
	13	**3** ·025^{+}	**2** ·007	**2** ·007	**1** ·001
	12	**2** ·016	**2** ·016	**1** ·003	**1** ·003
	11	**2** ·033	**1** ·008	**1** ·008	**0** ·001
	10	**1** ·017	**1** ·017	**0** ·002	**0** ·002
	9	**1** ·034	**0** ·006	**0** ·006	—
	8	**0** ·012	**0** ·012	—	—
	7	**0** ·024	**0** ·024	—	—
	6	**0** ·045^{+}	—	—	—

The table shows:

(1) In bold type, for given A, B and a, the value of b ($<a$) which is just significant at the probability level quoted (single-tail test).

(2) In small type, for given A, B and $r=a+b$, the exact probability (if there is independence) that b is equal to or less than the integer shown in bold type.

Significance tests in a 2 × 2 *contingency table* (*continued*)

	a	Probability			
		0·05	0·025	0·01	0·005
A = 16 B = 8	16	**5** ·028	**4** ·007	**4** ·007	**3** ·001
	15	**4** ·028	**3** ·007	**3** ·007	**2** ·001
	14	**3** ·021	**3** ·021	**2** ·005−	**2** ·005−
	13	**3** ·047	**2** ·013	**1** ·002	**1** ·002
	12	**2** ·028	**1** ·006	**1** ·006	**0** ·001
	11	**1** ·014	**1** ·014	**0** ·002	**0** ·002
	10	**1** ·027	**0** ·004	**0** ·004	**0** ·004
	9	**0** ·009	**0** ·009	**0** ·009	—
	8	**0** ·017	**0** ·017	—	—
	7	**0** ·033	—	—	—
7	16	**4** ·020	**4** ·020	**3** ·004	**3** ·004
	15	**3** ·017	**3** ·017	**2** ·003	**2** ·003
	14	**3** ·045+	**2** ·011	**1** ·002	**1** ·002
	13	**2** ·026	**1** ·005−	**1** ·005−	**1** ·005−
	12	**1** ·012	**1** ·012	**0** ·001	**0** ·001
	11	**1** ·024	**1** ·024	**0** ·003	**0** ·003
	10	**1** ·045−	**0** ·007	**0** ·007	—
	9	**0** ·014	**0** ·014	—	—
	8	**0** ·026	—	—	—
	7	**0** ·047	—	—	—
6	16	**3** ·013	**3** ·013	**2** ·002	**2** ·002
	15	**3** ·046	**2** ·009	**2** ·009	**1** ·001
	14	**2** ·025+	**1** ·004	**1** ·004	**1** ·004
	13	**1** ·011	**1** ·011	**0** ·001	**0** ·001
	12	**1** ·023	**1** ·023	**0** ·003	**0** ·003
	11	**1** ·043	**0** ·006	**0** ·006	—
	10	**0** ·012	**0** ·012	—	—
	9	**0** ·023	**0** ·023	—	—
	8	**0** ·040	—	—	—
5	16	**3** ·048	**2** ·008	**2** ·008	**1** ·001
	15	**2** ·028	**1** ·004	**1** ·004	**1** ·004
	14	**1** ·011	**1** ·011	**0** ·001	**0** ·001
	13	**1** ·025+	**0** ·003	**0** ·003	**0** ·003
	12	**1** ·047	**0** ·006	**0** ·006	—
	11	**0** ·012	**0** ·012	—	—
	10	**0** ·023	**0** ·023	—	—
	9	**0** ·039	—	—	—
4	16	**2** ·032	**1** ·004	**1** ·004	**1** ·004
	15	**1** ·013	**1** ·013	**0** ·001	**1** ·001
	14	**1** ·032	**0** ·003	**0** ·003	**0** ·003
	13	**0** ·007	**0** ·007	**0** ·007	—
	12	**0** ·014	**0** ·014	—	—
	11	**0** ·026	—	—	—
	10	**0** ·043	—	—	—
A = 16 B = 3	16	**1** ·018	**1** ·018	**0** ·001	**0** ·001
	15	**0** ·004	**0** ·004	**0** ·004	**0** ·004
	14	**0** ·010+	**0** ·010+	—	—
	13	**0** ·021	**0** ·021	—	—
	12	**0** ·036	—	—	—
2	16	**0** ·007	**0** ·007	**0** ·007	—
	15	**0** ·020	**0** ·020	—	—
	14	**0** ·039	—	—	—
A = 17 B = 17	17	**12** ·022	**12** ·022	**11** ·009	**10** ·004
	16	**11** ·043	**10** ·020	**9** ·008	**8** ·003
	15	**9** ·029	**8** ·013	**7** ·005+	**6** ·002
	14	**8** ·035+	**7** ·016	**6** ·007	**5** ·002
	13	**7** ·040	**6** ·018	**5** ·007	**4** ·003
	12	**6** ·042	**5** ·019	**4** ·007	**3** ·002
	11	**5** ·042	**4** ·018	**3** ·007	**2** ·002
	10	**4** ·040	**3** ·016	**2** ·005+	**1** ·001
	9	**3** ·035+	**2** ·013	**1** ·003	**1** ·003
	8	**2** ·029	**1** ·008	**1** ·008	**0** ·001
	7	**1** ·020	**1** ·020	**0** ·004	**0** ·004
	6	**1** ·043	**0** ·009	**0** ·009	—
	5	**0** ·022	**0** ·022	—	—
16	17	**12** ·044	**11** ·018	**10** ·007	**9** ·003
	16	**10** ·035−	**9** ·015−	**8** ·006	**7** ·002
	15	**9** ·046	**8** ·021	**7** ·009	**6** ·003
	14	**7** ·025+	**6** ·011	**5** ·004	**5** ·004
	13	**6** ·027	**5** ·011	**4** ·004	**4** ·004
	12	**5** ·027	**4** ·011	**3** ·004	**3** ·004
	11	**4** ·025+	**3** ·009	**3** ·009	**2** ·003
	10	**3** ·022	**3** ·022	**2** ·007	**1** ·002
	9	**3** ·046	**2** ·017	**1** ·004	**1** ·004
	8	**2** ·036	**1** ·011	**0** ·002	**0** ·002
	7	**1** ·024	**1** ·024	**0** ·005−	**0** ·005−
	6	**0** ·011	**0** ·011	—	—
	5	**0** ·026	—	—	—
15	17	**11** ·038	**10** ·015−	**9** ·006	**8** ·002
	16	**9** ·027	**8** ·011	**7** ·004	**7** ·004
	15	**8** ·035+	**7** ·015−	**6** ·006	**5** ·002
	14	**7** ·040	**6** ·017	**5** ·006	**4** ·002
	13	**6** ·041	**5** ·017	**4** ·006	**3** ·002
	12	**5** ·039	**4** ·016	**3** ·005+	**2** ·001
	11	**4** ·035+	**3** ·013	**2** ·004	**2** ·004
	10	**3** ·029	**2** ·010−	**2** ·010−	**1** ·002
	9	**2** ·022	**2** ·022	**1** ·006	**0** ·001
	8	**2** ·046	**1** ·014	**0** ·002	**0** ·002
	7	**1** ·030	**0** ·006	**0** ·006	—
	6	**0** ·014	**0** ·014	—	—
	5	**0** ·031	—	—	—

Significance tests in a 2 × 2 contingency table (continued)

	a	Probability			
		0·05	0·025	0·01	0·005
A=17 B=14	17	**10** ·032	**9** ·012	**8** ·004	**8** ·004
	16	**8** ·021	**8** ·021	**7** ·008	**6** ·003
	15	**7** ·026	**6** ·010−	**6** ·010−	**5** ·003
	14	**6** ·028	**5** ·011	**4** ·004	**4** ·004
	13	**5** ·027	**4** ·010−	**4** ·010−	**3** ·003
	12	**4** ·024	**4** ·024	**3** ·008	**2** ·002
	11	**4** ·049	**3** ·019	**2** ·006	**1** ·001
	10	**3** ·040	**2** ·014	**1** ·003	**1** ·003
	9	**2** ·029	**1** ·008	**1** ·008	**0** ·001
	8	**1** ·018	**1** ·018	**0** ·003	**0** ·003
	7	**1** ·038	**0** ·007	**0** ·007	—
	6	**0** ·017	**0** ·017	—	—
	5	**0** ·036	—	—	—
13	17	**9** ·026	**8** ·009	**8** ·009	**7** ·003
	16	**8** ·040	**7** ·015+	**6** ·005+	**5** ·002
	15	**7** ·045+	**6** ·018	**5** ·006	**4** ·002
	14	**6** ·045+	**5** ·018	**4** ·006	**3** ·002
	13	**5** ·042	**4** ·016	**3** ·005+	**2** ·001
	12	**4** ·035+	**3** ·013	**2** ·004	**2** ·004
	11	**3** ·028	**2** ·009	**2** ·009	**1** ·002
	10	**2** ·019	**2** ·019	**1** ·005−	**1** ·005−
	9	**2** ·040	**1** ·011	**0** ·002	**0** ·002
	8	**1** ·024	**1** ·024	**0** ·004	**0** ·004
	7	**1** ·047	**0** ·010−	**0** ·010−	—
	6	**0** ·021	**0** ·021	—	—
	5	**0** ·043	—	—	—
12	17	**8** ·021	**8** ·021	**7** ·007	**6** ·002
	16	**7** ·030	**6** ·011	**5** ·003	**5** ·003
	15	**6** ·033	**5** ·012	**4** ·004	**4** ·004
	14	**5** ·030	**4** ·011	**3** ·003	**3** ·003
	13	**4** ·026	**3** ·008	**3** ·008	**2** ·002
	12	**3** ·020	**3** ·020	**2** ·006	**1** ·001
	11	**3** ·041	**2** ·013	**1** ·003	**1** ·003
	10	**2** ·028	**1** ·007	**1** ·007	**0** ·001
	9	**1** ·016	**1** ·016	**0** ·002	**0** ·002
	8	**1** ·032	**0** ·006	**0** ·006	—
	7	**0** ·012	**0** ·012	—	—
	6	**0** ·026	—	—	—
11	17	**7** ·016	**7** ·016	**6** ·005−	**6** ·005−
	16	**6** ·022	**6** ·022	**5** ·007	**4** ·002
	15	**5** ·022	**5** ·022	**4** ·007	**3** ·002
	14	**4** ·019	**4** ·019	**3** ·006	**2** ·001
A=17 B=11	13	**4** ·042	**3** ·014	**2** ·004	**2** ·004
	12	**3** ·031	**2** ·009	**2** ·009	**1** ·002
	11	**2** ·020	**2** ·020	**1** ·005−	**1** ·005−
	10	**2** ·040	**1** ·011	**0** ·001	**0** ·001
	9	**1** ·022	**1** ·022	**0** ·004	**0** ·004
	8	**1** ·042	**0** ·008	**0** ·008	—
	7	**0** ·016	**0** ·016	—	—
	6	**0** ·033	—	—	—
10	17	**7** ·041	**6** ·012	**5** ·003	**5** ·003
	16	**6** ·047	**5** ·015+	**4** ·004	**4** ·004
	15	**5** ·043	**4** ·014	**3** ·004	**3** ·004
	14	**4** ·034	**3** ·010+	**2** ·002	**2** ·002
	13	**3** ·024	**3** ·024	**2** ·007	**1** ·001
	12	**3** ·049	**2** ·015+	**1** ·003	**1** ·003
	11	**2** ·031	**1** ·007	**1** ·007	**0** ·001
	10	**1** ·016	**1** ·016	**0** ·002	**0** ·002
	9	**1** ·031	**0** ·005+	**0** ·005+	—
	8	**0** ·011	**0** ·011	—	—
	7	**0** ·022	**0** ·022	—	—
	6	**0** ·042	—	—	—
9	17	**6** ·032	**5** ·008	**5** ·008	**4** ·002
	16	**5** ·034	**4** ·010−	**4** ·010−	**3** ·002
	15	**4** ·028	**3** ·008	**3** ·008	**2** ·002
	14	**3** ·020	**3** ·020	**2** ·005−	**2** ·005−
	13	**3** ·042	**2** ·012	**1** ·002	**1** ·002
	12	**2** ·025+	**1** ·006	**1** ·006	**0** ·001
	11	**2** ·048	**1** ·012	**0** ·002	**0** ·002
	10	**1** ·024	**1** ·024	**0** ·004	**0** ·004
	9	**1** ·045−	**0** ·008	**0** ·008	—
	8	**0** ·016	**0** ·016	—	—
	7	**0** ·030	—	—	—
8	17	**5** ·024	**5** ·024	**4** ·006	**3** ·001
	16	**4** ·023	**4** ·023	**3** ·006	**2** ·001
	15	**3** ·017	**3** ·017	**2** ·004	**2** ·004
	14	**3** ·039	**2** ·010−	**2** ·010−	**1** ·002
	13	**2** ·022	**2** ·022	**1** ·004	**1** ·004
	12	**2** ·043	**1** ·010−	**1** ·010−	**0** ·001
	11	**1** ·020	**1** ·020	**0** ·003	**0** ·003
	10	**1** ·038	**0** ·006	**0** ·006	—
	9	**0** ·012	**0** ·012	—	—
	8	**0** ·022	**0** ·022	—	—
	7	**0** ·040	—	—	—

The table shows:

(1) In bold type, for given A, B and a, the value of b $(<a)$ which is just significant at the probability level quoted (single-tail test).

(2) In small type, for given A, B and $r=a+b$, the exact probability (if there is independence) that b is equal to or less than the integer shown in bold type.

	a	Probability			
		0·05	0·025	0·01	0·005
A=17 B=7	17	**4** ·017	**4** ·017	**3** ·003	**3** ·003
	16	**3** ·014	**3** ·014	**2** ·003	**2** ·003
	15	**3** ·038	**2** ·009	**2** ·009	**1** ·001
	14	**2** ·021	**2** ·021	**1** ·004	**1** ·004
	13	**2** ·042	**1** ·009	**1** ·009	**0** ·001
	12	**1** ·018	**1** ·018	**0** ·002	**0** ·002
	11	**1** ·034	**0** ·005^{-}	**0** ·005^{-}	**0** ·005^{-}
	10	**0** ·010^{-}	**0** ·010^{-}	**0** ·010^{-}	—
	9	**0** ·019	**0** ·019	—	—
	8	**0** ·033	—	—	—
6	17	**3** ·011	**3** ·011	**2** ·002	**2** ·002
	16	**3** ·040	**2** ·008	**2** ·008	**1** ·001
	15	**2** ·021	**2** ·021	**1** ·003	**1** ·003
	14	**2** ·045^{+}	**1** ·009	**1** ·009	**0** ·001
	13	**1** ·018	**1** ·018	**0** ·002	**0** ·002
	12	**1** ·035^{-}	**0** ·005^{-}	**0** ·005^{-}	**0** ·005^{-}
	11	**0** ·009	**0** ·009	**0** ·009	—
	10	**0** ·017	**0** ·017	—	—
	9	**0** ·030	—	—	—
	8	**0** ·050^{-}	—	—	—
5	17	**3** ·043	**2** ·006	**2** ·006	**1** ·001
	16	**2** ·024	**2** ·024	**1** ·003	**1** ·003
	15	**1** ·009	**1** ·009	**1** ·009	**0** ·001
	14	**1** ·021	**1** ·021	**0** ·002	**0** ·002
	13	**1** ·039	**0** ·005^{-}	**0** ·005^{-}	**0** ·005^{-}
	12	**0** ·010^{-}	**0** ·010^{-}	**0** ·010^{-}	—
	11	**0** ·018	**0** ·018	—	—
	10	**0** ·030	—	—	—
	9	**0** ·049	—	—	—
4	17	**2** ·029	**1** ·003	**1** ·003	**1** ·003
	16	**1** ·011	**1** ·011	**0** ·001	**0** ·001
	15	**1** ·028	**0** ·003	**0** ·003	**0** ·003
	14	**0** ·006	**0** ·006	**0** ·006	—
	13	**0** ·012	**0** ·012	—	—
	12	**0** ·021	**0** ·021	—	—
	11	**0** ·035^{+}	—	—	—
3	17	**1** ·016	**1** ·016	**0** ·001	**0** ·001
	16	**1** ·046	**0** ·004	**0** ·004	**0** ·004
	15	**0** ·009	**0** ·009	**0** ·009	—
	14	**0** ·018	**0** ·018	—	—
	13	**0** ·031	—	—	—
	12	**0** ·049	—	—	—
2	17	**0** ·006	**0** ·006	**0** ·006	—
	16	**0** ·018	**0** ·018	—	—
	15	**0** ·035^{+}	—	—	—

	a	Probability			
		0·05	0·025	0·01	0·005
A=18 B=18	18	**13** ·023	**13** ·023	**12** ·010^{-}	**11** ·004
	17	**12** ·044	**11** ·020	**10** ·009	**9** ·004
	16	**10** ·030	**9** ·014	**8** ·006	**7** ·002
	15	**9** ·038	**8** ·018	**7** ·008	**6** ·003
	14	**8** ·043	**7** ·020	**6** ·009	**5** ·003
	13	**7** ·046	**6** ·022	**5** ·009	**4** ·003
	12	**6** ·047	**5** ·022	**4** ·009	**3** ·003
	11	**5** ·046	**4** ·020	**3** ·008	**2** ·002
	10	**4** ·043	**3** ·018	**2** ·006	**1** ·001
	9	**3** ·038	**2** ·014	**1** ·004	**1** ·004
	8	**2** ·030	**1** ·009	**1** ·009	**0** ·001
	7	**1** ·020	**1** ·020	**0** ·004	**0** ·004
	6	**1** ·044	**0** ·010^{-}	**0** ·010^{-}	—
	5	**0** ·023	**0** ·023	—	—
17	18	**13** ·045^{+}	**12** ·019	**11** ·008	**10** ·003
	17	**11** ·036	**10** ·016	**9** ·007	**8** ·002
	16	**10** ·049	**9** ·023	**8** ·010^{-}	**7** ·004
	15	**8** ·028	**7** ·012	**6** ·005^{-}	**6** ·005^{-}
	14	**7** ·030	**6** ·013	**5** ·005^{+}	**4** ·002
	13	**6** ·031	**5** ·013	**4** ·005^{-}	**4** ·005^{-}
	12	**5** ·030	**4** ·012	**3** ·004	**3** ·004
	11	**4** ·028	**3** ·010^{+}	**2** ·003	**2** ·003
	10	**3** ·023	**3** ·023	**2** ·008	**1** ·002
	9	**3** ·047	**2** ·018	**1** ·005^{-}	**1** ·005^{-}
	8	**2** ·037	**1** ·011	**0** ·002	**0** ·002
	7	**1** ·025^{-}	**1** ·025^{-}	**0** ·005^{-}	**0** ·005^{-}
	6	**0** ·011	**0** ·011	—	—
	5	**0** ·026	—	—	—
16	18	**12** ·039	**11** ·016	**10** ·006	**9** ·002
	17	**10** ·029	**9** ·012	**8** ·005^{-}	**8** ·005^{-}
	16	**9** ·038	**8** ·017	**7** ·007	**6** ·002
	15	**8** ·043	**7** ·019	**6** ·008	**5** ·003
	14	**7** ·046	**6** ·020	**5** ·008	**4** ·003
	13	**6** ·045^{+}	**5** ·020	**4** ·007	**3** ·002
	12	**5** ·042	**4** ·018	**3** ·006	**2** ·002
	11	**4** ·037	**3** ·015^{-}	**2** ·004	**2** ·004
	10	**3** ·031	**2** ·011	**1** ·003	**1** ·003
	9	**2** ·023	**2** ·023	**1** ·006	**0** ·001
	8	**2** ·046	**1** ·014	**0** ·002	**0** ·002
	7	**1** ·030	**0** ·006	**0** ·006	—
	6	**0** ·014	**0** ·014	—	—
	5	**0** ·031	—	—	—
15	18	**11** ·033	**10** ·013	**9** ·005^{-}	**9** ·005^{-}
	17	**9** ·023	**9** ·023	**8** ·009	**7** ·003
	16	**8** ·029	**7** ·012	**6** ·004	**6** ·004
	15	**7** ·031	**6** ·013	**5** ·005^{-}	**5** ·005^{-}
	14	**6** ·031	**5** ·013	**4** ·004	**4** ·004
	13	**5** ·029	**4** ·011	**3** ·004	**3** ·004

Significance tests in a 2 × 2 contingency table (continued)

	a	Probability			
		0·05	0·025	0·01	0·005
A=18 B=15	12	**4** ·025+	**3** ·009	**3** ·009	**2** ·003
	11	**3** ·020	**3** ·020	**2** ·006	**1** ·001
	10	**3** ·041	**2** ·014	**1** ·004	**1** ·004
	9	**2** ·030	**1** ·008	**1** ·008	**0** ·001
	8	**1** ·018	**1** ·018	**0** ·003	**0** ·003
	7	**1** ·038	**0** ·007	**0** ·007	—
	6	**0** ·017	**0** ·017	—	—
	5	**0** ·036	—	—	—
14	18	**10** ·028	**9** ·010−	**9** ·010−	**8** ·003
	17	**9** ·043	**8** ·017	**7** ·006	**6** ·002
	16	**8** ·050−	**7** ·021	**6** ·008	**5** ·003
	15	**6** ·022	**6** ·022	**5** ·008	**4** ·003
	14	**6** ·049	**5** ·020	**4** ·007	**3** ·002
	13	**5** ·044	**4** ·017	**3** ·006	**2** ·001
	12	**4** ·037	**3** ·013	**2** ·004	**2** ·004
	11	**3** ·028	**2** ·009	**2** ·009	**1** ·002
	10	**2** ·020	**2** ·020	**1** ·005−	**1** ·005−
	9	**2** ·039	**1** ·011	**0** ·002	**0** ·002
	8	**1** ·024	**1** ·024	**0** ·004	**0** ·004
	7	**1** ·047	**0** ·009	**0** ·009	—
	6	**0** ·020	**0** ·020	—	—
	5	**0** ·043	—	—	—
13	18	**9** ·023	**9** ·023	**8** ·008	**7** ·002
	17	**8** ·034	**7** ·012	**6** ·004	**6** ·004
	16	**7** ·037	**6** ·014	**5** ·005−	**5** ·005−
	15	**6** ·036	**5** ·014	**4** ·004	**4** ·004
	14	**5** ·032	**4** ·012	**3** ·004	**3** ·004
	13	**4** ·027	**3** ·009	**3** ·009	**2** ·002
	12	**3** ·020	**3** ·020	**2** ·006	**1** ·001
	11	**3** ·040	**2** ·013	**1** ·003	**1** ·003
	10	**2** ·027	**1** ·007	**1** ·007	**0** ·001
	9	**1** ·015+	**1** ·015+	**0** ·002	**0** ·002
	8	**1** ·031	**0** ·006	**0** ·006	—
	7	**0** ·012	**0** ·012	—	—
	6	**0** ·025+	—	—	—
12	18	**8** ·018	**8** ·018	**7** ·006	**6** ·002
	17	**7** ·026	**6** ·009	**6** ·009	**5** ·003
	16	**6** ·027	**5** ·009	**5** ·009	**4** ·003
	15	**5** ·024	**5** ·024	**4** ·008	**3** ·002
	14	**4** ·020	**4** ·020	**3** ·006	**2** ·001
	13	**4** ·042	**3** ·014	**2** ·004	**2** ·004
	12	**3** ·030	**2** ·009	**2** ·009	**1** ·002
	11	**2** ·019	**2** ·019	**1** ·005−	**1** ·005−
A=18 B=12	10	**2** ·038	**1** ·010+	**0** ·001	**0** ·001
	9	**1** ·021	**1** ·021	**0** ·003	**0** ·003
	8	**1** ·040	**0** ·007	**0** ·007	—
	7	**0** ·016	**0** ·016	—	—
	6	**0** ·031	—	—	—
11	18	**8** ·045+	**7** ·014	**6** ·004	**6** ·004
	17	**6** ·018	**6** ·018	**5** ·006	**4** ·001
	16	**5** ·018	**5** ·018	**4** ·005+	**3** ·001
	15	**5** ·043	**4** ·015−	**3** ·004	**3** ·004
	14	**4** ·033	**3** ·011	**2** ·003	**2** ·003
	13	**3** ·023	**3** ·023	**2** ·007	**1** ·001
	12	**3** ·046	**2** ·014	**1** ·003	**1** ·003
	11	**2** ·029	**1** ·007	**1** ·007	**0** ·001
	10	**1** ·015−	**1** ·015−	**0** ·002	**0** ·002
	9	**1** ·029	**0** ·005−	**0** ·005−	**0** ·005−
	8	**0** ·010+	**0** ·010+	—	—
	7	**0** ·020	**0** ·020	—	—
	6	**0** ·039	—	—	—
10	18	**7** ·037	**6** ·010+	**5** ·003	**5** ·003
	17	**6** ·041	**5** ·013	**4** ·003	**4** ·003
	16	**5** ·036	**4** ·011	**3** ·003	**3** ·003
	15	**4** ·028	**3** ·008	**3** ·008	**2** ·002
	14	**3** ·019	**3** ·019	**2** ·005−	**2** ·005−
	13	**3** ·039	**2** ·011	**1** ·002	**1** ·002
	12	**2** ·023	**2** ·023	**1** ·005+	**0** ·001
	11	**2** ·043	**1** ·011	**0** ·001	**0** ·001
	10	**1** ·022	**1** ·022	**0** ·003	**0** ·003
	9	**1** ·040	**0** ·007	**0** ·007	—
	8	**0** ·014	**0** ·014	—	—
	7	**0** ·027	—	—	—
	6	**0** ·049	—	—	—
9	18	**6** ·029	**5** ·007	**5** ·007	**4** ·002
	17	**5** ·030	**4** ·008	**4** ·008	**3** ·002
	16	**4** 023	**4** ·023	**3** ·006	**2** ·001
	15	**3** ·016	**3** ·016	**2** ·004	**2** ·004
	14	**3** ·034	**2** ·009	**2** ·009	**1** ·002
	13	**2** ·019	**2** ·019	**1** ·004	**1** ·004
	12	**2** ·037	**1** ·009	**1** ·009	**0** ·001
	11	**1** ·018	**1** ·018	**0** ·002	**0** ·002
	10	**1** ·033	**0** ·005+	**0** ·005+	—
	9	**0** ·010+	**0** ·010+	—	—
	8	**0** ·020	**0** ·020	—	—
	7	**0** ·036	—	—	—

The table shows:

(1) In bold type, for given A, B and a, the value of b ($<a$) which is just significant at the probability level quoted (single-tail test).

(2) In small type, for given A, B and $r=a+b$, the exact probability (if there is independence) that b is equal to or less than the integer shown in bold type.

	a	Probability			
		0·05	0·025	0·01	0·005
A = 18 B = 8	18	**5** ·022	**5** ·022	**4** ·005^{-}	**4** ·005^{-}
	17	**4** ·020	**4** ·020	**3** ·004	**3** ·004
	16	**3** ·014	**3** ·014	**2** ·003	**2** ·003
	15	**3** ·032	**2** ·008	**2** ·008	**1** ·001
	14	**2** ·017	**2** ·017	**1** ·003	**1** ·003
	13	**2** ·034	**1** ·007	**1** ·007	**0** ·001
	12	**1** ·015^{+}	**1** ·015^{+}	**0** ·002	**0** ·002
	11	**1** ·028	**0** ·004	**0** ·004	**0** ·004
	10	**1** ·049	**0** ·008	**0** ·008	—
	9	**0** ·016	**0** ·016	—	—
	8	**0** ·028	—	—	—
	7	**0** ·048	—	—	—
7	18	**4** ·015^{+}	**4** ·015^{+}	**3** ·003	**3** ·003
	17	**3** ·012	**3** ·012	**2** ·002	**2** ·002
	16	**3** ·032	**2** ·007	**2** ·007	**1** ·001
	15	**2** ·017	**2** ·017	**1** ·003	**1** ·003
	14	**2** ·034	**1** ·007	**1** ·007	**0** ·001
	13	**1** ·014	**1** ·014	**0** ·002	**0** ·002
	12	**1** ·027	**0** ·004	**0** ·004	**0** ·004
	11	**1** ·046	**0** ·007	**0** ·007	—
	10	**0** ·013	**0** ·013	—	—
	9	**0** ·024	**0** ·024	—	—
	8	**0** ·040	—	—	—
6	18	**3** ·010^{-}	**3** ·010^{-}	**3** ·010^{-}	**2** ·001
	17	**3** ·035^{+}	**2** ·006	**2** ·006	**1** ·001
	16	**2** ·018	**2** ·018	**1** ·003	**1** ·003
	15	**2** ·038	**1** ·007	**1** ·007	**0** ·001
	14	**1** ·015^{-}	**1** ·015^{-}	**0** ·002	**0** ·002
	13	**1** ·028	**0** ·003	**0** ·003	**0** ·003
	12	**1** ·048	**0** ·007	**0** ·007	—
	11	**0** ·013	**0** ·013	—	—
	10	**0** ·022	**0** ·022	—	—
	9	**0** ·037	—	—	—
5	18	**3** ·040	**2** ·006	**2** ·006	**1** ·001
	17	**2** ·021	**2** ·021	**1** ·003	**1** ·003
	16	**2** ·048	**1** ·008	**1** ·008	**0** ·001
	15	**1** ·017	**1** ·017	**0** ·002	**0** ·002
	14	**1** ·033	**0** ·004	**0** ·004	**0** ·004
	13	**0** ·007	**0** ·007	**0** ·007	—
	12	**0** ·014	**0** ·014	—	—
	11	**0** ·024	**0** ·024	—	—
	10	**0** ·038	—	—	—
4	18	**2** ·026	**1** ·003	**1** ·003	**1** ·003
	17	**1** ·010^{-}	**1** ·010^{-}	**1** ·010^{-}	**0** ·001
	16	**1** ·024	**1** ·024	**0** ·002	**0** ·002
	15	**1** ·046	**0** ·005^{-}	**0** ·005^{-}	**0** ·005^{-}
	14	**0** ·010^{-}	**0** ·010^{-}	**0** ·010^{-}	—
A = 18 B = 4	13	**0** ·017	**0** ·017	—	—
	12	**0** ·029	—	—	—
	11	**0** ·045^{+}	—	—	—
3	18	**1** ·014	**1** ·014	**0** ·001	**0** ·001
	17	**1** ·041	**0** ·003	**0** ·003	**0** ·003
	16	**0** ·008	**0** ·008	**0** ·008	—
	15	**0** ·015^{+}	**0** ·015^{+}	—	—
	14	**0** ·026	—	—	—
	13	**0** ·042	—	—	—
2	18	**0** ·005^{+}	**0** ·005^{+}	**0** ·005^{+}	—
	17	**0** ·016	**0** ·016	—	—
	16	**0** ·032	—	—	—
A = 19 B = 19	19	**14** ·023	**14** ·023	**13** ·010^{-}	**12** ·004
	18	**13** ·045^{-}	**12** ·021	**11** ·009	**10** ·004
	17	**11** ·031	**10** ·015^{-}	**9** ·006	**8** ·003
	16	**10** ·039	**9** ·019	**8** ·009	**7** ·003
	15	**9** ·046	**8** ·022	**6** ·004	**6** ·004
	14	**8** ·050^{-}	**7** ·024	**5** ·004	**5** ·004
	13	**6** ·025^{+}	**5** ·011	**4** ·004	**4** ·004
	12	**5** ·024	**5** ·024	**3** ·003	**3** ·003
	11	**5** ·050^{-}	**4** ·022	**3** ·009	**2** ·003
	10	**4** ·046	**3** ·019	**2** ·006	**1** ·002
	9	**3** ·039	**2** ·015^{-}	**1** ·004	**1** ·004
	8	**2** ·031	**1** ·009	**1** ·009	**0** ·002
	7	**1** ·021	**1** ·021	**0** ·004	**0** ·004
	6	**1** ·045^{-}	**0** ·010^{-}	**0** ·010^{-}	—
	5	**0** ·023	**0** ·023	—	—
18	19	**14** ·046	**13** ·020	**12** ·008	**11** ·003
	18	**12** ·037	**11** ·017	**10** ·007	**9** ·003
	17	**10** ·024	**10** ·024	**8** ·004	**8** ·004
	16	**9** ·030	**8** ·014	**7** ·006	**6** ·002
	15	**8** ·033	**7** ·015^{+}	**6** ·006	**5** ·002
	14	**7** ·035^{+}	**6** ·016	**5** ·006	**4** ·002
	13	**6** ·035^{-}	**5** ·015^{+}	**4** ·006	**3** ·002
	12	**5** ·033	**4** ·014	**3** ·005^{-}	**3** ·005^{-}
	11	**4** ·030	**3** ·011	**2** ·004	**2** ·004
	10	**3** ·025^{-}	**3** ·025^{-}	**2** ·008	**1** ·002
	9	**3** ·049	**2** ·019	**1** ·005^{+}	**0** ·001
	8	**2** ·038	**1** ·012	**0** ·002	**0** ·002
	7	**1** ·025^{+}	**0** ·005^{-}	**0** ·005^{-}	**0** ·005^{-}
	6	**0** ·012	**0** ·012	—	—
	5	**0** ·027	—	—	—
17	19	**13** ·040	**12** ·016	**11** ·006	**10** ·002
	18	**11** ·030	**10** ·013	**9** ·005^{+}	**8** ·002
	17	**10** ·040	**9** ·018	**8** ·008	**7** ·003
	16	**9** ·047	**8** ·022	**7** ·009	**6** ·003

	a	Probability			
		0·05	0·025	0·01	0·005
A = 19 B = 17	15	**8** ·050−	**7** ·023	**6** ·010−	**5** ·004
	14	**6** ·023	**6** ·023	**5** ·010−	**4** ·003
	13	**6** ·049	**5** ·022	**4** ·008	**3** ·003
	12	**5** ·045−	**4** ·019	**3** ·007	**2** ·002
	11	**4** ·039	**3** ·015+	**2** ·005−	**2** ·005−
	10	**3** ·032	**2** ·011	**1** ·003	**1** ·003
	9	**2** ·024	**2** ·024	**1** ·007	**0** ·001
	8	**2** ·047	**1** ·015−	**0** ·002	**0** ·002
	7	**1** ·031	**0** ·006	**0** ·006	—
	6	**0** ·014	**0** ·014	—	—
	5	**0** ·031	—	—	—
16	19	**12** ·035−	**11** ·013	**10** ·005−	**10** ·005−
	18	**10** ·024	**10** ·024	**9** ·010−	**8** ·004
	17	**9** ·031	**8** ·013	**7** ·005+	**6** ·002
	16	**8** ·035−	**7** ·015+	**6** ·006	**5** ·002
	15	**7** ·036	**6** ·015+	**5** ·006	**4** ·002
	14	**6** ·034	**5** ·014	**4** ·005+	**3** ·002
	13	**5** ·031	**4** ·013	**3** ·004	**3** ·004
	12	**4** ·027	**3** ·010−	**3** ·010−	**2** ·003
	11	**3** ·021	**3** ·021	**2** ·007	**1** ·002
	10	**3** ·042	**2** ·015−	**1** ·004	**1** ·004
	9	**2** ·030	**1** ·009	**1** ·009	**0** ·001
	8	**1** ·018	**1** ·018	**0** ·003	**0** ·003
	7	**1** ·037	**0** ·007	**0** ·007	—
	6	**0** ·017	**0** ·017	—	—
	5	**0** ·036	—	—	—
15	19	**11** ·029	**10** ·011	**9** ·004	**9** ·004
	18	**10** ·046	**9** ·019	**8** ·007	**7** ·002
	17	**8** ·023	**8** ·023	**7** ·009	**6** ·003
	16	**7** ·025−	**7** ·025−	**6** ·010−	**5** ·003
	15	**6** ·024	**6** ·024	**5** ·009	**4** ·003
	14	**5** ·022	**5** ·022	**4** ·008	**3** ·002
	13	**5** ·045+	**4** ·018	**3** ·006	**2** ·002
	12	**4** ·037	**3** ·014	**2** ·004	**2** ·004
	11	**3** ·029	**2** ·009	**2** ·009	**1** ·002
	10	**2** ·020	**2** ·020	**1** ·005+	**0** ·001
	9	**2** ·039	**1** ·011	**0** ·002	**0** ·002
	8	**1** ·023	**1** ·023	**0** ·004	**0** ·004
	7	**1** ·046	**0** ·009	**0** ·009	—
	6	**0** ·020	**0** ·020	—	—
	5	**0** ·042	—	—	—
14	19	**10** ·024	**10** ·024	**9** ·008	**8** ·003
	18	**9** ·037	**8** ·014	**7** ·005−	**7** ·005−
	17	**8** ·042	**7** ·017	**6** ·006	**5** ·002
A = 19 B = 14	16	**7** ·042	**6** ·017	**5** ·006	**4** ·002
	15	**6** ·039	**5** ·015+	**4** ·005+	**3** ·001
	14	**5** ·034	**4** ·013	**3** ·004	**3** ·004
	13	**4** ·027	**3** ·009	**3** ·009	**2** ·003
	12	**3** ·020	**3** ·020	**2** ·006	**1** ·001
	11	**3** ·040	**2** ·013	**1** ·003	**1** ·003
	10	**2** ·027	**1** ·007	**1** ·007	**0** ·001
	9	**1** ·015−	**1** ·015−	**0** ·002	**0** ·002
	8	**1** ·030	**0** ·005+	**0** ·005+	—
	7	**0** ·012	**0** ·012	—	—
	6	**0** ·024	**0** ·024	—	—
	5	**0** ·049	—	—	—
13	19	**9** ·020	**9** ·020	**8** ·006	**7** ·002
	18	**8** ·029	**7** ·010+	**6** ·003	**6** ·003
	17	**7** ·031	**6** ·011	**5** ·004	**5** ·004
	16	**6** ·029	**5** ·011	**4** ·003	**4** ·003
	15	**5** ·025+	**4** ·009	**4** ·009	**3** ·003
	14	**4** ·020	**4** ·020	**3** ·006	**2** ·002
	13	**4** ·041	**3** ·015−	**2** ·004	**2** ·004
	12	**3** ·029	**2** ·009	**2** ·009	**1** ·002
	11	**2** ·019	**2** ·019	**1** ·005−	**1** ·005−
	10	**2** ·036	**1** ·010−	**1** ·010−	**0** ·001
	9	**1** ·020	**1** ·020	**0** ·003	**0** ·003
	8	**1** ·038	**0** ·007	**0** ·007	—
	7	**0** ·015−	**0** ·015−	—	—
	6	**0** ·030	—	—	—
12	19	**9** ·049	**8** ·016	**7** ·005−	**7** ·005−
	18	**7** ·022	**7** ·022	**6** ·007	**5** ·002
	17	**6** ·022	**6** ·022	**5** ·007	**4** ·002
	16	**5** ·019	**5** ·019	**4** ·006	**3** ·002
	15	**5** ·042	**4** ·015+	**3** ·004	**3** ·004
	14	**4** ·032	**3** ·011	**2** ·003	**2** ·003
	13	**3** ·023	**3** ·023	**2** ·006	**1** ·001
	12	**3** ·043	**2** ·014	**1** ·003	**1** ·003
	11	**2** ·027	**1** ·007	**1** ·007	**0** ·001
	10	**2** ·050−	**1** ·014	**0** ·002	**0** ·002
	9	**1** ·027	**0** ·005−	**0** ·005−	**0** ·005−
	8	**1** ·050−	**0** ·010−	**0** ·010−	—
	7	**0** ·019	**0** ·019	—	—
	6	**0** ·037	—	—	—
11	19	**8** ·041	**7** ·012	**6** ·003	**6** ·003
	18	**7** ·047	**6** ·016	**5** ·004	**5** ·004
	17	**6** ·043	**5** ·015−	**4** ·004	**4** ·004
	16	**5** ·035+	**4** ·012	**3** ·003	**3** ·003

The table shows:

(1) In bold type, for given *A*, *B* and *a*, the value of *b* ($< a$) which is just significant at the probability level quoted (single-tail test).

(2) In small type, for given *A*, *B* and $r = a + b$, the exact probability (if there is independence) that *b* is equal to or less than the integer shown in bold type.

	a	Probability			
		0·05	0·025	0·01	0·005
A = 19 B = 11	15	4 ·027	3 ·008	3 ·008	2 ·002
	14	3 ·018	3 ·018	2 ·005^{-}	2 ·005^{-}
	13	3 ·035^{+}	2 ·010^{+}	1 ·002	1 ·002
	12	2 ·021	2 ·021	1 ·005^{-}	1 ·005^{-}
	11	2 ·040	1 ·010^{+}	0 ·001	0 ·001
	10	1 ·020	1 ·020	0 ·003	0 ·003
	9	1 ·037	0 ·006	0 ·006	—
	8	0 ·013	0 ·013	—	—
	7	0 ·025^{-}	0 ·025^{-}	—	—
	6	0 ·046	—	—	—
10	19	7 ·033	6 ·009	6 ·009	5 ·002
	18	6 ·036	5 ·011	4 ·003	4 ·003
	17	5 ·030	4 ·009	4 ·009	3 ·002
	16	4 ·022	4 ·022	3 ·006	2 ·001
	15	4 ·047	3 ·015^{-}	2 ·004	2 ·004
	14	3 ·030	2 ·008	2 ·008	1 ·002
	13	2 ·017	2 ·017	1 ·004	1 ·004
	12	2 ·033	1 ·008	1 ·008	0 ·001
	11	1 ·016	1 ·016	0 ·002	0 ·002
	10	1 ·029	0 ·005^{-}	0 ·005^{-}	0 ·005^{-}
	9	0 ·009	0 ·009	0 ·009	—
	8	0 ·018	0 ·018	—	—
	7	0 ·032	—	—	—
9	19	6 ·026	5 ·006	5 ·006	4 ·001
	18	5 ·026	4 ·007	4 ·007	3 ·001
	17	4 ·020	4 ·020	3 ·005^{-}	3 ·005^{-}
	16	4 ·044	3 ·013	2 ·003	2 ·003
	15	3 ·028	2 ·007	2 ·007	1 ·001
	14	2 ·015^{-}	2 ·015^{-}	1 ·003	1 ·003
	13	2 ·029	1 ·006	1 ·006	0 ·001
	12	1 ·013	1 ·013	0 ·002	0 ·002
	11	1 ·024	1 ·024	0 ·004	0 ·004
	10	1 ·042	0 ·007	0 ·007	—
	9	0 ·013	0 ·013	—	—
	8	0 ·024	0 ·024	—	—
	7	0 ·043	—	—	—
8	19	5 ·019	5 ·019	4 ·004	4 ·004
	18	4 ·017	4 ·017	3 ·004	3 ·004
	17	4 ·044	3 ·011	2 ·002	2 ·002
	16	3 ·027	2 ·006	2 ·006	1 ·001
	15	2 ·013	2 ·013	1 ·002	1 ·002
	14	2 ·027	1 ·006	1 ·006	0 ·001
	13	2 ·049	1 ·011	0 ·001	0 ·001
	12	1 ·021	1 ·021	0 ·003	0 ·003
	11	1 ·038	0 ·006	0 ·006	—
	10	0 ·011	0 ·011	—	—
	9	0 ·020	0 ·020	—	—
	8	0 ·034	—	—	—
A = 19 B = 7	19	4 ·013	4 ·013	3 ·002	3 ·002
	18	4 ·047	3 ·010^{+}	2 ·002	2 ·002
	17	3 ·028	2 ·006	2 ·006	1 ·001
	16	2 ·014	2 ·014	1 ·002	1 ·002
	15	2 ·028	1 ·005^{+}	1 ·005^{+}	0 ·001
	14	1 ·011	1 ·011	0 ·001	0 ·001
	13	1 ·021	1 ·021	0 ·003	0 ·003
	12	1 ·037	0 ·005^{+}	0 ·005^{+}	—
	11	0 ·010^{-}	0 ·010^{-}	0 ·010^{-}	—
	10	0 ·017	0 ·017	—	—
	9	0 ·030	—	—	—
	8	0 ·048	—	—	—
6	19	4 ·050^{-}	3 ·009	3 ·009	2 ·001
	18	3 ·031	2 ·005^{+}	2 ·005^{+}	1 ·001
	17	2 ·015^{+}	2 ·015^{+}	1 ·002	1 ·002
	16	2 ·032	1 ·006	1 ·006	0 ·000
	15	1 ·012	1 ·012	0 ·001	0 ·001
	14	1 ·023	1 ·023	0 ·003	0 ·003
	13	1 ·039	0 ·005^{+}	0 ·005^{+}	—
	12	0 ·010^{-}	0 ·010^{-}	0 ·010^{-}	—
	11	0 ·017	0 ·017	—	—
	10	0 ·028	—	—	—
	9	0 ·045^{+}	—	—	—
5	19	3 ·036	2 ·005^{-}	2 ·005^{-}	2 ·005^{-}
	18	2 ·018	2 ·018	1 ·002	1 ·002
	17	2 ·042	1 ·006	1 ·006	0 ·000
	16	1 ·014	1 ·014	0 ·001	0 ·001
	15	1 ·028	0 ·003	0 ·003	0 ·003
	14	1 ·047	0 ·006	0 ·006	—
	13	0 ·011	0 ·011	—	—
	12	0 ·019	0 ·019	—	—
	11	0 ·030	—	—	—
	10	0 ·047	—	—	—
4	19	2 ·024	2 ·024	1 ·002	1 ·002
	18	1 ·009	1 ·009	1 ·009	0 ·001
	17	1 ·021	1 ·021	0 ·002	0 ·002
	16	1 ·040	0 ·004	0 ·004	0 ·004
	15	0 ·008	0 ·008	0 ·008	—
	14	0 ·014	0 ·014	—	—
	13	0 ·024	0 ·024	—	—
	12	0 ·037	—	—	—
3	19	1 ·013	1 ·013	0 ·001	0 ·001
	18	1 ·038	0 ·003	0 ·003	0 ·003
	17	0 ·006	0 ·006	0 ·006	—
	16	0 ·013	0 ·013	—	—
	15	0 ·023	0 ·023	—	—
	14	0 ·036	—	—	—

	a	Probability			
		0·05	0·025	0·01	0·005
A=19 B=2	19	**0** ·005−	**0** ·005−	**0** ·005−	**0** ·005−
	18	**0** ·014	**0** ·014	—	—
	17	**0** ·029	—	—	—
	16	**0** ·048	—	—	—
A=20 B=20	20	**15** ·024	**15** ·024	**13** ·004	**13** ·004
	19	**14** ·046	**13** ·022	**12** ·010−	**11** ·004
	18	**12** ·032	**11** ·015+	**10** ·007	**9** ·003
	17	**11** ·041	**10** ·020	**9** ·009	**8** ·004
	16	**10** ·048	**9** ·024	**7** ·005−	**7** ·005−
	15	**8** ·027	**7** ·012	**6** ·005+	**5** ·002
	14	**7** ·028	**6** ·013	**5** ·005+	**4** ·002
	13	**6** ·028	**5** ·012	**4** ·005−	**4** ·005−
	12	**5** ·027	**4** ·011	**3** ·004	**3** ·004
	11	**4** ·024	**4** ·024	**3** ·009	**2** ·003
	10	**4** ·048	**3** ·020	**2** ·007	**1** ·002
	9	**3** ·041	**2** ·015+	**1** ·004	**1** ·004
	8	**2** ·032	**1** ·010−	**1** ·010−	**0** ·002
	7	**1** ·022	**1** ·022	**0** ·004	**0** ·004
	6	**1** ·046	**0** ·010+	—	—
	5	**0** ·024	**0** ·024	—	—
19	20	**15** ·047	**14** ·020	**13** ·008	**12** ·003
	19	**13** ·039	**12** ·018	**11** ·008	**10** ·003
	18	**11** ·026	**10** ·012	**9** ·005−	**9** ·005−
	17	**10** ·032	**9** ·015−	**8** ·006	**7** ·002
	16	**9** ·036	**8** ·017	**7** ·007	**6** ·003
	15	**8** ·038	**7** ·018	**6** ·008	**5** ·003
	14	**7** ·039	**6** ·018	**5** ·007	**4** ·003
	13	**6** ·038	**5** ·017	**4** ·007	**3** ·002
	12	**5** ·035+	**4** ·015+	**3** ·005+	**2** ·002
	11	**4** ·031	**3** ·012	**2** ·004	**2** ·004
	10	**3** ·026	**2** ·009	**2** ·009	**1** ·002
	9	**2** ·019	**2** ·019	**1** ·005+	**0** ·001
	8	**2** ·039	**1** ·012	**0** ·002	**0** ·002
	7	**1** ·026	**0** ·005+	**0** ·005+	—
	6	**0** ·012	**0** ·012	—	—
	5	**0** ·027	—	—	—
18	20	**14** ·041	**13** ·017	**12** ·007	**11** ·003
	19	**12** ·032	**11** ·014	**10** ·006	**9** ·002
	18	**11** ·043	**10** ·020	**9** ·008	**8** ·003
	17	**10** ·050−	**9** ·024	**7** ·004	**7** ·004
	16	**8** ·026	**7** ·011	**6** ·005−	**6** ·005−
A=20 B=18	15	**7** ·027	**6** ·012	**5** ·004	**5** ·004
	14	**6** ·026	**5** ·011	**4** ·004	**4** ·004
	13	**5** ·024	**5** ·024	**4** ·009	**3** ·003
	12	**5** ·047	**4** ·020	**3** ·007	**2** ·002
	11	**4** ·041	**3** ·016	**2** ·005+	**1** ·001
	10	**3** ·033	**2** ·012	**1** ·003	**1** ·003
	9	**2** ·024	**2** ·024	**1** ·007	**0** ·001
	8	**2** ·048	**1** ·015−	**0** ·003	**0** ·003
	7	**1** ·031	**0** ·006	**0** ·006	—
	6	**0** ·014	**0** ·014	—	—
	5	**0** ·031	—	—	—
17	20	**13** ·036	**12** ·014	**11** ·005+	**10** ·002
	19	**11** ·026	**10** ·011	**9** ·004	**9** ·004
	18	**10** ·034	**9** ·015−	**8** ·006	**7** ·002
	17	**9** ·038	**8** ·017	**7** ·007	**6** ·003
	16	**8** ·040	**7** ·018	**6** ·007	**5** ·003
	15	**7** ·039	**6** ·017	**5** ·007	**4** ·002
	14	**6** ·037	**5** ·016	**4** ·006	**3** ·002
	13	**5** ·033	**4** ·013	**3** ·005−	**3** ·005−
	12	**4** ·028	**3** ·010+	**2** ·003	**2** ·003
	11	**3** ·022	**3** ·022	**2** ·007	**1** ·002
	10	**3** ·042	**2** ·015+	**1** ·004	**1** ·004
	9	**2** ·031	**1** ·009	**1** ·009	**0** ·001
	8	**1** ·019	**1** ·019	**0** ·003	**0** ·003
	7	**1** ·037	**0** ·008	**0** ·008	—
	6	**0** ·017	**0** ·017	—	—
	5	**0** ·036	—	—	—
16	20	**12** ·031	**11** ·012	**10** ·004	**10** ·004
	19	**11** ·049	**10** ·021	**9** ·008	**8** ·003
	18	**9** ·026	**8** ·011	**7** ·004	**7** ·004
	17	**8** ·028	**7** ·012	**6** ·004	**6** ·004
	16	**7** ·028	**6** ·012	**5** ·004	**5** ·004
	15	**6** ·026	**5** ·011	**4** ·004	**4** ·004
	14	**5** ·023	**5** ·023	**4** ·009	**3** ·003
	13	**5** ·046	**4** ·019	**3** ·007	**2** ·002
	12	**4** ·038	**3** ·014	**2** ·004	**2** ·004
	11	**3** ·029	**2** ·010−	**2** ·010−	**1** ·002
	10	**2** ·020	**2** ·020	**1** ·005+	**0** ·001
	9	**2** ·039	**1** ·011	**0** ·002	**0** ·002
	8	**1** ·023	**1** ·023	**0** ·004	**0** ·004
	7	**1** ·045+	**0** ·009	**0** ·009	—
	6	**0** ·020	**0** ·020	—	—
	5	**0** ·041	—	—	—

The table shows:

(1) In bold type, for given A, B and a, the value of b ($<a$) which is just significant at the probability level quoted (single-tail test).

(2) In small type, for given A, B and $r=a+b$, the exact probability (if there is independence) that b is equal t or less than the integer shown in bold type.

	a	Probability			
		0·05	0·025	0·01	0·005
A=20 B=15	20	11 ·026	10 ·009	10 ·009	9 ·003
	19	10 ·040	9 ·016	8 ·006	7 ·002
	18	9 ·046	8 ·019	7 ·007	6 ·002
	17	8 ·047	7 ·020	6 ·008	5 ·002
	16	7 ·045−	6 ·019	5 ·007	4 ·002
	15	6 ·040	5 ·017	4 ·006	3 ·002
	14	5 ·034	4 ·013	3 ·004	3 ·004
	13	4 ·028	3 ·010−	3 ·010−	2 ·003
	12	3 ·020	3 ·020	2 ·006	1 ·001
	11	3 ·039	2 ·013	1 ·003	1 ·003
	10	2 ·026	1 ·007	1 ·007	0 ·001
	9	2 ·049	1 ·015−	0 ·002	0 ·002
	8	1 ·029	0 ·005+	0 ·005+	—
	7	0 ·012	0 ·012	—	—
	6	0 ·024	0 ·024	—	—
	5	0 ·048	—	—	—
14	20	10 ·022	10 ·022	9 ·007	8 ·002
	19	9 ·032	8 ·012	7 ·004	7 ·004
	18	8 ·035+	7 ·014	6 ·005−	6 ·005−
	17	7 ·035−	6 ·013	5 ·005−	5 ·005−
	16	6 ·031	5 ·012	4 ·004	4 ·004
	15	5 ·026	4 ·009	4 ·009	3 ·003
	14	4 ·020	4 ·020	3 ·007	2 ·002
	13	4 ·040	3 ·015−	2 ·004	2 ·004
	12	3 ·029	2 ·009	2 ·009	1 ·002
	11	2 ·018	2 ·018	1 ·005−	1 ·005−
	10	2 ·035+	1 ·010−	1 ·010−	0 ·001
	9	1 ·019	1 ·019	0 ·003	0 ·003
	8	1 ·037	0 ·007	0 ·007	—
	7	0 ·014	0 ·014	—	—
	6	0 ·029	—	—	—
13	20	9 ·017	9 ·017	8 ·005+	7 ·002
	19	8 ·025−	8 ·025−	7 ·008	6 ·003
	18	7 ·026	6 ·009	6 ·009	5 ·003
	17	6 ·024	6 ·024	5 ·008	4 ·002
	16	5 ·020	5 ·020	4 ·007	3 ·002
	15	5 ·041	4 ·015+	3 ·005−	3 ·005−
	14	4 ·031	3 ·011	2 ·003	2 ·003
	13	3 ·022	3 ·022	2 ·006	1 ·001
	12	3 ·041	2 ·013	1 ·003	1 ·003
	11	2 ·026	1 ·007	1 ·007	0 ·001
	10	2 ·047	1 ·013	0 ·002	0 ·002
	9	1 ·026	0 ·004	0 ·004	0 ·004
	8	1 ·047	0 ·009	0 ·009	—
	7	0 ·018	0 ·018	—	—
	6	0 ·035−	—	—	—
12	20	9 ·044	8 ·014	7 ·004	7 ·004
	19	7 ·019	7 ·019	6 ·006	5 ·002
A=20 B=12	18	6 ·018	6 ·018	5 ·006	4 ·002
	17	6 ·043	5 ·016	4 ·005−	4 ·005−
	16	5 ·034	4 ·012	3 ·003	3 ·003
	15	4 ·025+	3 ·008	3 ·008	2 ·002
	14	4 ·049	3 ·017	2 ·005−	2 ·005−
	13	3 ·033	2 ·010−	2 ·010−	1 ·002
	12	2 ·020	2 ·020	1 ·005−	1 ·005−
	11	2 ·036	1 ·009	1 ·009	0 ·001
	10	1 ·018	1 ·018	0 ·003	0 ·003
	9	1 ·034	0 ·006	0 ·006	—
	8	0 ·012	0 ·012	—	—
	7	0 ·023	0 ·023	—	—
	6	0 ·043	—	—	—
11	20	8 ·037	7 ·010+	6 ·003	6 ·003
	19	7 ·042	6 ·013	5 ·004	5 ·004
	18	6 ·037	5 ·012	4 ·003	4 ·003
	17	5 ·029	4 ·009	4 ·009	3 ·002
	16	4 ·021	4 ·021	3 ·006	2 ·001
	15	4 ·042	3 ·014	2 ·003	2 ·003
	14	3 ·028	2 ·008	2 ·008	1 ·001
	13	2 ·016	2 ·016	1 ·003	1 ·003
	12	2 ·029	1 ·007	1 ·007	0 ·001
	11	1 ·014	1 ·014	0 ·002	0 ·002
	10	1 ·026	0 ·004	0 ·004	0 ·004
	9	1 ·046	0 ·008	0 ·008	—
	8	0 ·016	0 ·016	—	—
	7	0 ·029	—	—	—
10	20	7 ·030	6 ·008	6 ·008	5 ·002
	19	6 ·031	5 ·009	5 ·009	4 ·002
	18	5 ·026	4 ·007	4 ·007	3 ·002
	17	4 ·018	4 ·018	3 ·005−	3 ·005−
	16	4 ·039	3 ·012	2 ·003	2 ·003
	15	3 ·024	3 ·024	2 ·006	1 ·001
	14	3 ·045+	2 ·013	1 ·003	1 ·003
	13	2 ·025+	1 ·006	1 ·006	0 ·001
	12	2 ·045−	1 ·011	0 ·001	0 ·001
	11	1 ·021	1 ·021	0 ·003	0 ·003
	10	1 ·037	0 ·006	0 ·006	—
	9	0 ·012	0 ·012	—	—
	8	0 ·022	0 ·022	—	—
	7	0 ·038	—	—	—
9	20	6 ·023	6 ·023	5 ·005+	4 ·001
	19	5 ·022	5 ·022	4 ·005+	3 ·001
	18	4 ·016	4 ·016	3 ·004	3 ·004
	17	4 ·037	3 ·010+	2 ·002	2 ·002
	16	3 ·022	3 ·022	2 ·005+	1 ·001
	15	3 ·043	2 ·012	1 ·002	1 ·002
	14	2 ·023	2 ·023	1 ·005−	1 ·005−

	a	Probability 0·05	0·025	0·01	0·005
A=20 B=9	13	**2** ·041	**1** ·009	**1** ·009	**0** ·001
	12	**1** ·018	**1** ·018	**0** ·002	**0** ·002
	11	**1** ·032	**0** ·005−	**0** ·005−	**0** ·005−
	10	**0** ·009	**0** ·009	**0** ·009	—
	9	**0** ·017	**0** ·017	—	—
	8	**0** ·029	—	—	—
	7	**0** ·050−	—	—	—
8	20	**5** ·017	**5** ·017	**4** ·003	**4** ·003
	19	**4** ·015−	**4** ·015−	**3** ·003	**3** ·003
	18	**4** ·038	**3** ·009	**3** ·009	**2** ·002
	17	**3** ·022	**3** ·022	**2** ·005−	**2** ·005−
	16	**3** ·044	**2** ·011	**1** ·002	**1** ·002
	15	**2** ·022	**2** ·022	**1** ·004	**1** ·004
	14	**2** ·040	**1** ·009	**1** ·009	**0** ·001
	13	**1** ·016	**1** ·016	**0** ·002	**0** ·002
	12	**1** ·029	**0** ·004	**0** ·004	**0** ·004
	11	**1** ·048	**0** ·008	**0** ·008	—
	10	**0** ·014	**0** ·014	—	—
	9	**0** ·024	**0** ·024	—	—
	8	**0** ·041	—	—	—
7	20	**4** ·012	**4** ·012	**3** ·002	**3** ·002
	19	**4** ·042	**3** ·009	**3** ·009	**2** ·001
	18	**3** ·024	**3** ·024	**2** ·005−	**2** ·005−
	17	**3** ·050−	**2** ·011	**1** ·002	**1** ·002
	16	**2** ·023	**2** ·023	**1** ·004	**1** ·004
	15	**2** ·043	**1** ·009	**1** ·009	**0** ·001
	14	**1** ·016	**1** ·016	**0** ·002	**0** ·002
	13	**1** ·029	**0** ·004	**0** ·004	**0** ·004
	12	**1** ·048	**0** ·007	**0** ·007	—
	11	**0** ·013	**0** ·013	—	—
	10	**0** ·022	**0** ·022	—	—
	9	**0** ·036	—	—	—
6	20	**4** ·046	**3** ·008	**3** ·008	**2** ·001
	19	**3** ·028	**2** ·005−	**2** ·005−	**2** ·005−
	18	**2** ·013	**2** ·013	**1** ·002	**1** ·002
	17	**2** ·028	**1** ·004	**1** ·004	**1** ·004
	16	**1** ·010−	**1** ·010−	**1** ·010−	**0** ·001
	15	**1** ·018	**1** ·018	**0** ·002	**0** ·002

	a	Probability 0·05	0·025	0·01	0·005
A=20 B=6	14	**1** ·032	**0** ·004	**0** ·004	**0** ·004
	13	**0** ·007	**0** ·007	**0** ·007	—
	12	**0** ·013	**0** ·013	—	—
	11	**0** ·022	**0** ·022	—	—
	10	**0** ·035−	—	—	—
5	20	**3** ·033	**2** ·004	**2** ·004	**2** ·004
	19	**2** ·016	**2** ·016	**1** ·002	**1** ·002
	18	**2** ·038	**1** ·005+	**1** ·005+	**0** ·000
	17	**1** ·012	**1** ·012	**0** ·001	**0** ·001
	16	**1** ·023	**1** ·023	**0** ·002	**0** ·002
	15	**1** ·040	**0** ·005−	**0** ·005−	**0** ·005−
	14	**0** ·009	**0** ·009	**0** ·009	—
	13	**0** ·015−	**0** ·015−	—	—
	12	**0** ·024	**0** ·024	—	—
	11	**0** ·038	—	—	—
4	20	**2** ·022	**2** ·022	**1** ·002	**1** ·002
	19	**1** ·008	**1** ·008	**1** ·008	**0** ·000
	18	**1** ·018	**1** ·018	**0** ·001	**0** ·001
	17	**1** ·035+	**0** ·003	**0** ·003	**0** ·003
	16	**0** ·007	**0** ·007	**0** ·007	—
	15	**0** ·012	**0** ·012	—	—
	14	**0** ·020	**0** ·020	—	—
	13	**0** ·031	—	—	—
	12	**0** ·047	—	—	—
3	20	**1** ·012	**1** ·012	**0** ·001	**0** ·001
	19	**1** ·034	**0** ·002	**0** ·002	**0** ·002
	18	**0** ·006	**0** ·006	**0** ·006	—
	17	**0** ·011	**0** ·011	—	—
	16	**0** ·020	**0** ·020	—	—
	15	**0** ·032	—	—	—
	14	**0** ·047	—	—	—
2	20	**0** ·004	**0** ·004	**0** ·004	**0** ·004
	19	**0** ·013	**0** ·013	—	—
	18	**0** ·026	—	—	—
	17	**0** ·043	—	—	—
1	20	**0** ·048	—	—	—

The table shows:

(1) In bold type, for given A, B and a, the value of b ($<a$) which is just significant at the probability lev quoted (single-tail test).

(2) In small type, for given A, B and $r=a+b$, the exact probability (if there is independence) that b is equal t or less than the integer shown in bold type.

A METHOD FOR JUDGING ALL CONTRASTS IN THE ANALYSIS OF VARIANCE*

BY HENRY SCHEFFÉ, *Columbia University*

A simple answer is found for the following question which has plagued the practice of the analysis of variance: Under the usual assumptions, if the conventional F-test of the hypothesis $H: \mu_1 = \mu_2 = \ldots = \mu_k$ at the α level of significance rejects H, what further inferences are valid about the contrasts among the μ_i (beyond the inference that the values of the contrasts are not all zero)? Suppose the F-test has $k-1$ and ν degrees of freedom. For any $c_1, \ldots, c_k$ with $\sum_1^k c_i = 0$ write θ for the contrast $\sum_1^k c_i \mu_i$, and write $\hat{\theta}$ and $\hat{\sigma}^2_{\hat{\theta}}$ for the usual estimates of θ and the variance of $\hat{\theta}$. Then for the totality of contrasts, no matter what the true values of the θ's, the probability is $1-\alpha$ that they all satisfy

$$\hat{\theta} - S\hat{\sigma}_{\hat{\theta}} \leqslant \theta \leqslant \hat{\theta} + S\hat{\sigma}_{\hat{\theta}},$$

where S^2 is $(k-1)$ times the upper α point of the F-distribution with $k-1$ and ν degrees of freedom. Suppose we say that the estimated contrast $\hat{\theta}$ is 'significantly different from zero' if $|\hat{\theta}| > S\hat{\sigma}_{\hat{\theta}}$. Then the F-test rejects H if and only if some $\hat{\theta}$ are significantly different from zero, and if it does, we can say just which $\hat{\theta}$. More generally, the above inequality can be employed for all the contrasts with the obvious frequency interpretation about the proportion of experiments in which all statements are correct. Relations are considered to an earlier method of Tukey using the Studentized range tables and valid in the special case where the $\hat{\mu}_i$ all have the same variance and all pairs $\hat{\mu}_i$, $\hat{\mu}_j$ $(i \neq j)$ have the same covariance. Some results are obtained for the operating characteristic of the new method. The paper is organized so that the reader who wishes to learn the method and avoid the proofs may skip §§ 2 and 5.

1. STATEMENT OF THE METHOD

The general problem is that of making inferences about the contrasts among a set of 'true means' or 'true main effects' $\mu_1, \mu_2, \ldots, \mu_k$ in the analysis of variance. For example, the μ_i might be the true row effects in a two-way lay-out with possibly unequal numbers of observations per cell. The μ_i may be unrestricted or subject to a single restriction of the form

$$\sum_1^k h_i \mu_i = h, \tag{1}$$

where the h_i and h are known constants with $\sum_1^k h_i \neq 0$. *A contrast* is a linear function of the μ_i,

$$\theta = \sum_1^k c_i \mu_i, \tag{2}$$

determined by k known constants c_i satisfying the condition

$$\sum_1^k c_i = 0. \tag{3}$$

The value of the linear function for a particular set of μ_i will be called the *value* of the contrast; it will not cause any confusion in the following to use the same symbol θ both for the contrast and the value of the contrast.

We make the assumptions usual in the analysis of variance, namely, that there is at hand a set of statistics $\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_k$, and $\hat{\sigma}^2$, such that the $\hat{\mu}_i$ have a multivariate normal distribution and are statistically independent of $\hat{\sigma}^2$, that

$$\begin{aligned} &E(\hat{\mu}_i) = \mu_i \quad (i = 1, \ldots, k), \\ \text{and} \quad &\operatorname{cov}(\hat{\mu}_i, \hat{\mu}_j) = a_{ij}\sigma^2 \quad (i,j = 1, \ldots, k), \end{aligned} \tag{4}$$

where the constants a_{ij} are known, and σ^2 is unknown. The μ_i will always be 'Model I' (non-random) effects, as discussed by Eisenhart (1947) or Mood (1950). In a pure Model I

---

* Work sponsored by the Office of Naval Research (U.S.A.).

situation, σ^2 is the variance σ_e^2 of a single observation ('error variance'). In a mixed model situation where there exists an exact F-test of the hypothesis

$$H\colon \mu_1 = \mu_2 = \ldots = \mu_k, \tag{5}$$

σ^2 will equal σ_e^2 plus further unknown non-negative parameters. In any case, $\hat{\sigma}^2$ is an estimate of σ^2 with ν D.F. (degrees of freedom), that is, $\nu\hat{\sigma}^2/\sigma^2$ has the χ^2 distribution with ν D.F. The case where σ^2 is known can be treated by obvious modifications of the theory below, usually merely by putting $\nu = \infty$ and $\hat{\sigma}^2 = \sigma^2$ in the results. It is further assumed that if the μ_i are unrestricted the rank* of the covariance matrix with elements (4) is k, and that if the μ_i are subject to a restriction (1) then the $\hat{\mu}_i$ are subject to the same restriction (1) and the rank of the covariance matrix is $k-1$.

The hypothesis H in (5), equivalent to the statement that all the contrasts are zero, can be tested by the conventional F-statistic with $k-1$ and ν D.F. We shall refer to this test at significance level α as 'the' F-test of H. The problem of making further inferences about the contrasts, arising when the F-test rejects H, has been considered by various writers, including R. A. Fisher (1935), D. Newman (1939), J. W. Tukey (1951), and H. K. Nandi (1951). Except for Tukey's and Nandi's, the methods involve repeated tests of significance on the same data, and are hence subject to the usual objection that little is known about the joint operating characteristic. While it is often not possible in practical applications to avoid repeated tests of significance, it is possible for the particular problem we are considering.

The solution studied in this paper is based on the following probability statement about the infinite totality of contrasts:† For any contrast (2) denote its estimate by $\hat{\theta}$,

$$\hat{\theta} = \sum_1^k c_i \hat{\mu}_i,$$

the variance of $\hat{\theta}$ by $\sigma_{\hat{\theta}}^2$,

$$\sigma_{\hat{\theta}}^2 = \sum_{i=1}^k \sum_{j=1}^k a_{ij} c_i c_j \sigma^2, \tag{6}$$

and the estimate of this variance by $\hat{\sigma}_{\hat{\theta}}^2$,

$$\hat{\sigma}_{\hat{\theta}}^2 = \sum_{i=1}^k \sum_{j=1}^k a_{ij} c_i c_j \hat{\sigma}^2.$$

Define the positive constant S from

$$S^2 = (k-1)\, F_\alpha(k-1, \nu), \tag{7}$$

---

[*] The non-mathematical statistician may safely assume these rank conditions to be satisfied in practical applications; they are stated because they are needed later for the mathematical arguments in §§2 and 5.

[†] The idea of making an overall confidence statement for all the contrasts, its successful realization, and the resultant possibility of making valid tests of hypotheses suggested by the data, I first met in a lecture by Prof. J. W. Tukey on 'New methods in the analysis of variance: Range in the numerator and range in the denominator' at the Annual Princeton Conference of the American Society for Quality Control on 8 December 1951. This method was published (Tukey, 1951). The method of Tukey described in §3 was explained in 'Allowances for various types of error rates', an unpublished invited address presented before a joint meeting of the Institute of Mathematical Statistics and the Eastern North American Region of the Biometric Society on 19 March 1952 at Blacksburg, Va.; it differs from the published method (Tukey, 1951) only in the use of a root-mean-square instead of a range estimate of the error standard deviation. Recently when I communicated the method based on (8) to Prof. Tukey he wrote me that it had been familiar to him and Prof. D. B. Duncan for some time, and that they had discussed it publicly when he gave a lecture at Blacksburg in November 1951.

where $F_\alpha(k-1,\nu)$ denotes the upper α point of the F-distribution with $k-1$ and ν D.F. Then the probability is $1-\alpha$ that the values θ of *all* the contrasts simultaneously satisfy

$$\hat{\theta}-S\hat{\sigma}_{\hat{\theta}} \leqslant \theta \leqslant \hat{\theta}+S\hat{\sigma}_{\hat{\theta}}, \tag{8}$$

no matter what the values of all unknown parameters. The proof of this statement will be given in §2.

This result may be used for the interval estimation of all contrasts of interest, including any suggested by the way the observed means $\hat{\mu}_i$ fall out. No matter how many contrasts are estimated by the method (8), the probability that all the statements thus made will be correct will be $\geqslant 1-\alpha$.

The result may also be used to declare any estimated contrast 'significantly different from zero' or not, according as the corresponding interval (8) excludes $\theta = 0$ or not. More precisely, after selecting a set of coefficients c_i subject to (3) and thus determining a contrast we make one of the following three statements (it will be convenient also to say we make the statement *for* the contrast as well as *about* its estimate):

(i) $\hat{\theta}$ *is not significantly different from zero,*
(ii) $\hat{\theta}$ *is significantly different from zero and positive,*
(iii) $\hat{\theta}$ *is significantly different from zero and negative.*

We make statement (i) if $-S\hat{\sigma}_{\hat{\theta}} < \hat{\theta} < S\hat{\sigma}_{\hat{\theta}}$, (ii) if $\hat{\theta} \geqslant S\hat{\sigma}_{\hat{\theta}}$, (iii) if $\hat{\theta} \leqslant -S\hat{\sigma}_{\hat{\theta}}$. The operating characteristic of this method is studied in §4.

We warn the reader here that in the special case where all the $\hat{\mu}_i$ have the same variance, and all pairs $\hat{\mu}_i$, $\hat{\mu}_j$ $(i \neq j)$ have the same covariance, and where the *only* contrasts of interest are the $\frac{1}{2}k(k-1)$ differences $\mu_i - \mu_j$, the method of Tukey described in §3 should be used in preference to the above, because the confidence intervals will then be shorter. An example of such a case may be found in Scheffé (1952).

2. PROOF OF THE METHOD

The proof will be made with the aid of a linear transformation and other mathematical apparatus which will again be useful in §5 for proving results about the operating characteristic stated in §4. Although the coefficients of the transformation will be regarded as known in the mathematical discussion, they need not be computed for the practical applications.

The dimension of the space of estimated contrasts $\hat{\theta}$, regarded as linear forms in indeterminates $\hat{\mu}_1$, is $k-1$. Under the assumptions we made about the rank of the covariance matrix (4), we may find a basis for this space such that the corresponding random variables $\hat{\eta}_1, \ldots, \hat{\eta}_{k-1}$, will be statistically independent with equal variance

$$\sigma^2_{\hat{\eta}} = C^2\sigma^2,$$

where C is some chosen positive constant. The choice of C for the purposes of §4 will be discussed there; it does not matter at present. Let $\hat{\eta}_k = k^{-\frac{1}{2}} \sum\limits_1^k \hat{\mu}_i$. Then between the $\hat{\mu}_i$ and the $\hat{\eta}_i$ there will exist a non-singular linear relationship,

$$\hat{\mu}_i = \sum_{j=1}^{k} b_{ij}\hat{\eta}_j \quad (i = 1, \ldots, k), \tag{9}$$

where the b_{ij} are constants not depending on unknown parameters. Writing $E(\hat{\eta}_j) = \eta_j$, we get from (9)

$$\mu_i = \sum_{j=1}^{k} b_{ij}\eta_j \quad (i = 1, \ldots, k).$$

Then

$$\theta = \sum_{i=1}^{k} c_i \mu_i = \sum_{i=1}^{k} \sum_{j=1}^{k} c_i b_{ij} \eta_j.$$

The coefficient of η_k must be zero for all $c_1, \ldots, c_k$ satisfying (3), so b_{ik} does not depend on i, and we may write

$$\theta = \sum_{j=1}^{k-1} d_j \eta_j\ \hat{\theta} = \sum_{j=1}^{k-1} d_j \hat{\eta}_j,$$

where

$$d_j = \sum_{i=1}^{k} c_i b_{ij} \quad (j = 1, \ldots, k-1).$$

Now

$$\sigma_{\hat{\theta}}^2 = \sum_{j=1}^{k-1} d_j^2 \sigma_{\hat{\eta}}^2 = C^2\sigma^2 \sum_{j=1}^{k-1} d_j^2. \tag{10}$$

A contrast θ for which $\sigma_{\hat{\theta}}^2 = C^2\sigma^2$ will be called a *normalized* contrast and denoted by ϑ. For a normalized contrast

$$\vartheta = \sum_{1}^{k} c_i \mu_i = \sum_{1}^{k-1} d_i \eta_i, \tag{11}$$

we then have

$$\sum_{1}^{k-1} d_i^2 = 1.$$

Clearly it suffices to prove that the probability statement associated with (8) holds for the totality of normalized contrasts. We shall do this by means of some simple geometric considerations.

Let us introduce a $(k-1)$-dimensional space, which we shall call the y-space of points $y = (y_1, \ldots, y_{k-1})$, for graphing the parameter point or vector $\eta = (\eta_1, \ldots, \eta_{k-1})$, its estimate $\hat{\eta}$, and other quantities. The random variables $\sum_{1}^{k-1} (\hat{\eta}_i - \eta_i)^2/(C^2\sigma^2)$ and $\nu\hat{\sigma}^2/\sigma^2$ have independent χ^2 distributions with $k-1$ and ν D.F., respectively, and so $\nu/(k-1)$ times their quotient has the F-distribution. This yields the confidence sphere $\mathscr{S}$,

$$\sum_{i=1}^{k-1} (y_i - \hat{\eta}_i)^2 \leqslant S^2 C^2 \hat{\sigma}^2, \tag{12}$$

for the parameter point η, where S^2 is defined by (7). The probability is $1-\alpha$ that the parameter point η is covered by $\mathscr{S}$.

Now a normalized contrast is uniquely determined by a coefficient vector $d = (d_1, \ldots, d_{k-1})$ of unit length, and it will be convenient in this section and § 5 to identify the contrast with the vector. It is seen from (11) that the value ϑ of the contrast is the projection of η on d. Similarly, the value $\hat{\vartheta}$ of its estimate is the projection of $\hat{\eta}$ on d. The interval (8) when written for ϑ becomes

$$\hat{\vartheta} - SC\hat{\sigma} \leqslant \vartheta \leqslant \hat{\vartheta} + SC\hat{\sigma}. \tag{13}$$

If we lay this interval off on the vector d, its centre is the projection on d of the centre $\hat{\eta}$ of the sphere $\mathscr{S}$, and its half-length is the radius of $\mathscr{S}$; in other words, the interval (13) may be interpreted as the projection of $\mathscr{S}$ on d. The interval covers the true value of the contrast if and only if the projection of the point η on d lies in the projection of $\mathscr{S}$ on d. This happens for all vectors d if and only if $\mathscr{S}$ covers η, and the probability of this is $1-\alpha$.

3. COMPARISON WITH A METHOD OF TUKEY

In the special case where all the $\hat{\mu}_i$ have the same variance $a_{11}\sigma^2$ and all pairs $\hat{\mu}_i, \hat{\mu}_j\ (i \neq j)$ have the same covariance $a_{12}\sigma^2$, the following method of Tukey* is applicable. The probability is $1-\alpha$ that the values θ of all the contrasts simultaneously satisfy

$$\hat{\theta} - T\hat{\sigma} \leqslant \theta \leqslant \hat{\theta} + T\hat{\sigma}, \tag{14}$$

where the constant T is defined as

$$T = \frac{1}{2}\sum_1^k |c_i|\, q(a_{11}-a_{12})^{\frac{1}{2}}, \tag{15}$$

and q is the upper α point of the Studentized range, for the range of a sample of k in the numerator, and ν D.F. in the denominator, that is, the upper α point of the quotient w/s, where w and s $(s > 0)$ are statistically independent, w is the range of a random sample of k standard normal deviates, and νs^2 has the χ^2 distribution with ν D.F. This has been tabled by J. M. May (1952) for $\alpha = 0{\cdot}05$ and $0{\cdot}01$, and needs to be tabled for $\alpha = 0{\cdot}10$.

We propose to compare the efficiency of the two methods by use of the ratio R of the squared lengths of the confidence intervals (8) and (14),

$$R = (S^2\hat{\sigma}^2_{\hat{\theta}})/(T^2\hat{\sigma}^2).$$

The motivation for using the squared lengths is that if for a particular contrast the value of R thus defined is R_0, we may say that for large samples the method (8) requires R_0 times as many measurements as (14) to give the same accuracy on this contrast.

After noting that

$$\sigma^2_{\hat{\theta}} = (a_{11}-a_{12})\,\sigma^2 \sum_1^k c_i^2$$

in the present case, we may express the ratio R as

$$R = \frac{S^2}{q^2}\left(\sum_1^k c_i^2\right)\Big/\left(\frac{1}{2}\sum_1^k |c_i|\right)^2. \tag{16}$$

The contrasts usually of the greatest practical interest are perhaps those consisting of the difference between the average of m of the μ_i and the average of r of the other μ_i $(m + r \leqslant k)$; we shall symbolize this type of contrast by $\{m, r\}$. For instance, a contrast of the type $\{2, 3\}$ is

$$\tfrac{1}{2}(\mu_2+\mu_7) - \tfrac{1}{3}(\mu_1+\mu_4+\mu_8). \tag{17}$$

It may be shown that R attains its maximum value $R_{\max.}$ for a contrast of the type $\{1, 1\}$, that is, a difference of two μ_i, and its minimum value $R_{\min.}$ for one of the type $\{\frac{1}{2}k, \frac{1}{2}k\}$ if k is even, $\{\frac{1}{2}(k-1), \frac{1}{2}(k+1)\}$ if k is odd, and hence

$$R_{\max.} = 2(S^2/q^2),$$

$$R_{\min.} = \begin{cases} 4k^{-1}(S^2/q^2) & \text{for } k \text{ even},\\ 4k(k^2-1)^{-1}(S^2/q^2) & \text{for } k \text{ odd}. \end{cases}$$

Table 1 shows how the relative efficiency of the two methods varies with k, the number of means, in the case $\nu = \infty$. The rows headed $1/R_{\max.}$ show the efficiency of method (8) relative to method (14) on contrasts which are differences of two μ_i, $R_{\min.}$ shows the efficiency of method (14) relative to method (8) on some other contrasts. The value of S^2/q^2 is also tabled for use with (16) for the calculation of R. Table 2 shows that the value of R is not very sensitive to ν. Of course any increase of S^2/q^2 above its value for $\nu = \infty$ listed in Table 1 favours method (14).

---

* See footnote on p. 88.

Table 1. *Variation of relative efficiency R of two methods for different contrasts* ($\nu = \infty$)

α	Value of	For $k =$ 2	3	4	5	6	8	10	13	16	20
0·10	S^2/q^2	0·50	0·55	0·60	0·64	0·69	0·78	0·86	0·98	1·09	1·23
	$1/R_{\text{max.}}$	1·00	0·91	0·84	0·78	0·73	0·64	0·58	0·51	0·46	0·40
	$R_{\text{min.}}$	1·00	0·82	0·60	0·54	0·46	0·39	0·34	0·30	0·27	0·25
0·05	S^2/q^2	0·50	0·54	0·59	0·64	0·68	0·76	0·85	0·96	1·07	1·20
	$1/R_{\text{max.}}$	1·00	0·92	0·84	0·79	0·73	0·65	0·59	0·52	0·47	0·42
	$R_{\text{min.}}$	1·00	0·82	0·59	0·53	0·45	0·38	0·34	0·30	0·27	0·24
0·01	S^2/q^2	0·50	0·54	0·59	0·63	0·67	0·74	0·81	0·92	1·01	1·14
	$1/R_{\text{max.}}$	1·00	0·92	0·85	0·80	0·75	0·67	0·61	0·55	0·49	0·44
	$R_{\text{min.}}$	1·00	0·81	0·59	0·52	0·44	0·37	0·33	0·28	0·25	0·23

Table 2. *Values of S^2/q^2, showing dependence of relative efficiency on ν*

α	ν \ k	3	6	10	13	16	20
0·05	5	0·56	0·70	0·89	1·03	1·15	1·31
	7	0·55	0·69	0·87	1·00	1·13	1·28
	10	0·55	0·69	0·87	0·99	1·11	1·26
	20	0·55	0·68	0·86	0·98	1·09	1·25
	40	0·55	0·68	0·85	0·97	1·08	1·23
	∞	0·54	0·68	0·85	0·96	1·07	1·20
0·01	5	0·57	0·72	0·91	1·05	1·18	1·34
	7	0·55	0·69	0·87	1·00	1·12	1·27
	10	0·55	0·68	0·85	0·98	1·09	1·24
	20	0·55	0·68	0·84	0·96	1·07	1·21
	40	0·54	0·67	0·83	0·94	1·04	1·18
	∞	0·54	0·67	0·81	0·92	1·01	1·14

The difference between the two methods is seen to increase with k; as k increases (8) gets relatively worse on the differences $\mu_i - \mu_j$, while (14) gets relatively worse on some other contrasts (and gets worse faster). It is clear that the choice between the two methods in the special case where (14) is applicable depends on which kind of contrast we think will be of interest.* (It is, of course, not permissible to use both on the same data and then choose the one with the results we like better—unless we are willing to settle for an overall confidence coefficient known only to be $\geqslant 1-2\alpha$, in which case we may choose for each contrast the shorter of the two intervals.) If we are interested exclusively in the differences $\mu_i-\mu_j$, we should choose (14); if we are interested in many types of contrasts, and in investigating contrasts suggested by the data, (8) would seem superior. Tables 3*a* and 3*b* show the relative efficiencies in the cases $k = 4$ and $k = 6$ on most contrasts likely to be of practical interest.* The type $\{m, r\}$ is described above (17). The contrasts for linear, quadratic, etc., effects are the contrasts used in fitting orthogonal polynomials when the μ_i correspond to equal steps of some independent variable. The coefficients c_i for this type of contrast are listed in Fisher & Yates's tables (1943). The relative efficiency of method (14) is given in the column headed R, of (8) in the column $1/R$. It is seen that for $k = 4$, method (14) is superior only on the type $\{1, 1\}$, for $k = 6$ also the type $\{1, 2\}$. These tables are for $\nu = \infty$ and require some correction in favour of (14) for small values of ν; the correction factor is the ratio of entries in Table 2 for the given ν and for $\nu = \infty$ at $\alpha = 0{\cdot}05$.

---

* For the reader who has not skipped §2, a further comparison of the two methods is contained in the following loose but perhaps helpful statements: In using method (8) the user 'pays' for more than he will use. If he knew beforehand just which contrasts were going to interest him, the corresponding tangent planes to the confidence sphere $\mathscr{S}$ of §2, normal to the coefficient vectors d of the contrasts, could be constructed, and these would bound a circumscribed polyhedron $\mathscr{H}$. The user pays for the information that the parameter point is in $\mathscr{S}$ but he uses only the information that it is in $\mathscr{H}$. In

Table 3*a*. *Relative efficiency of two methods when* $k = 4$ $(\alpha = 0{\cdot}05, \nu = \infty)$

Type of contrast	$1/R$	R
$\{1, 1\}$	0·84	
$\{1, 2\}$		0·89
$\{1, 3\}$		0·79
$\{2, 2\}$, quadratic		0·59
Linear, cubic		0·74

Table 3*b*. *Relative efficiency when* $k = 6$ $(\alpha = 0{\cdot}05, \nu = \infty)$

Type of contrast	$1/R$	R
$\{1, 1\}$	0·73	
$\{1, 2\}$	0·98	
$\{1, 3\}$		0·91
$\{1, 4\}$		0·85
$\{1, 5\}$		0·82
$\{2, 2\}$		0·68
$\{2, 3\}$		0·57
$\{2, 4\}$		0·51
$\{3, 3\}$		0·45
Linear		0·59
Quadratic		0·57
Cubic		0·48
Quartic		0·53
Quintic		0·67

principle it is possible to calculate an $\mathscr{H}'$ obtained by a uniform contraction of $\mathscr{H}$ about the centre of $\mathscr{S}$ and to pay just for this; in practice this would be a hopelessly complicated calculation for most cases. In the special case where the $\hat{\mu}_i$ have equal variances and equal covariances, and the contrasts of interest are the $\frac{1}{2}k(k-1)$ differences $\mu_i-\mu_j$, Tukey's method does calculate the $\mathscr{H}'$. If $\mathscr{H}'$ is available, we can of course infer about other contrasts by projecting $\mathscr{H}'$ on their vectors d, as we projected $\mathscr{S}$. This is equivalent to (14) for Tukey's method, and we will see below that this does not work as well for many contrasts of interest. We suggest that it will often be worthwhile to the customer to pay for more than he will use if he does not know beforehand just what he will need and can buy a good package.

* Interactions may also be regarded as contrasts. For example, if μ_{ij} is the true mean for the (i, j) cell in a two-way lay-out, the two-factor interaction in the (1, 2) cell is $\mu_{12}-\mu_{1.}-\mu_{.2}+\mu_{..}$, where the dots have their usual connotation, and this is a contrast among the μ_{ij}. General superiority of (8) over (14) for the interaction contrasts is indicated by some numerical calculations.

4. OPERATING CHARACTERISTIC OF THE METHOD

For many purposes it may suffice to decide whether an experiment has adequate sensitivity by considering the lengths of the confidence intervals (8). In this section we consider the effects of the method from a viewpoint close to the Neyman-Pearson (1933) concept of the two kinds of error possible in hypothesis-testing. We preface this rather long development with a few remarks.

The power of the method will turn out to seem low to one accustomed to power calculations for a single t-test.* This has its counterpart in estimation, in that for $k > 2$ our confidence interval (8) for a contrast will usually be much longer than the $100(1-\alpha)$ % confidence interval for this contrast alone based on the t-distribution (which it is, of course, not valid to apply if the choice of the contrast has been suggested by the configuration of the observed means). The ratio of the squared lengths of the confidence intervals, namely, t^2/S^2, where t is the upper $\frac{1}{2}\alpha$ point of the t-distribution with ν D.F. and S^2 is given by (7), is listed in Table 4. While these ratios depart far from unity for $k > 2$, the writer was somewhat surprised that the departure was not greater. For $k = 6$, for example, it requires only about three times as many measurements to get confidence intervals for as many contrasts as we please, including any suggested by the data, with 95 % confidence that all are correct, as for a 95 % confidence interval for a single one of the contrasts selected before the data are examined, the confidence interval in both cases having the same expected length.

Table 4. *Values of t^2/S^2*

α	k ν	2	3	4	5	6	8	10	13	16	20
0·10	5	1·00	0·54	0·37	0·29	0·24	0·17	0·14	0·10	0·08	0·07
	7	1·00	0·55	0·39	0·30	0·25	0·18	0·15	0·11	0·09	0·07
	10	1·00	0·56	0·40	0·32	0·26	0·19	0·16	0·12	0·10	0·08
	20	1·00	0·57	0·42	0·33	0·28	0·21	0·17	0·13	0·11	0·09
	40	1·00	0·58	0·42	0·34	0·28	0·22	0·18	0·14	0·11	0·09
	∞	1·00	0·59	0·43	0·35	0·29	0·23	0·18	0·15	0·12	0·10
0·05	5	1·00	0·57	0·41	0·32	0·26	0·19	0·15	0·12	0·10	0·08
	7	1·00	0·59	0·43	0·34	0·28	0·21	0·17	0·13	0·11	0·09
	10	1·00	0·61	0·45	0·36	0·30	0·23	0·18	0·14	0·12	0·09
	20	1·00	0·62	0·47	0·38	0·32	0·25	0·20	0·16	0·13	0·11
	40	1·00	0·63	0·48	0·39	0·33	0·26	0·21	0·17	0·14	0·12
	∞	1·00	0·64	0·49	0·40	0·35	0·27	0·23	0·18	0·15	0·13
0·01	5	1·00	0·61	0·45	0·36	0·30	0·22	0·18	0·14	0·11	0·09
	7	1·00	0·64	0·48	0·39	0·33	0·25	0·20	0·16	0·13	0·10
	10	1·00	0·66	0·51	0·42	0·36	0·28	0·23	0·18	0·15	0·12
	20	1·00	0·69	0·55	0·46	0·39	0·31	0·26	0·21	0·18	0·14
	40	1·00	0·71	0·57	0·48	0·42	0·33	0·28	0·23	0·19	0·16
	∞	1·00	0·72	0·58	0·50	0·44	0·36	0·31	0·25	0·22	0·18

---

* Those accustomed in applied statistics to making repeated t-tests at the 5 % significance level might consider choosing $\alpha = 10$ % rather than 5 % with the new method (8): The user of the repeated 5 % tests is working at some 'overall' significance level that is unknown but greater than 5 %; perhaps he would be glad to settle for a guaranteed 10 %. How bad the repeated t method may get to be from the 'overall' point of view is indicated below for confidence intervals in connexion with Tables 5 and 6.

Instead of considering the present method from the point of view of one accustomed to the method of repeated t-tests, or repeated confidence intervals based on the same data, it is instructive also to do the reverse. The theory of the method (8) permits the calculation of a lower bound for the overall confidence coefficient implied by the use of repeated t confidence intervals for the contrasts, calculated from the same data. Thus, for repeated 95 % confidence intervals based on t we may calculate the bound by substituting in place of S in (7) the two-tailed 5 % point of t with ν D.F., and solving the resulting equation for $1-\alpha$. The values of this bound shown in Table 5 may startle the reader when he considers that by increasing the number of repeated t confidence intervals, the overall probability may be brought arbitrarily close to this bound. If repeated t confidence intervals are used only on the differences $\mu_i-\mu_j$, and if the $\hat{\mu}_i$ have equal variances and equal covariances, then the theory of Tukey's method (14) leads to a better bound, found by solving for $1-\alpha$ the equation $q = 2^{\frac{1}{2}}t_{0\cdot05}$, where q is the same function of α, k, ν as in (15). This bound is attained if all $\frac{1}{2}k(k-1)$ statements are made about the differences $\mu_i-\mu_j$. Values of the bound are shown in Table 6. We may try to take comfort from the thought that if we always made all $\frac{1}{2}k(k-1)$ statements based on 5 % t, we should in the long run of experiments have 95 % of all the statements we made correct, but we should remember that of the statements to which we are likely to pay the most attention more than 5 % tend to be wrong; for example, statements associated with the larger observed differences.

Table 5. *Lower bound for probability that all repeated t confidence intervals for contrasts will be correct* ($\nu = \infty$)

Conf. coeff. \ k	2	3	4	5	6	8	10	13	16	20
0·90	0·90	0·74	0·56	0·39	0·25	0·09	0·03	0·003	$0{\cdot}0^{3}2$	$<0{\cdot}0^{4}1$
0·95	0·95	0·85	0·72	0·57	0·43	0·20	0·08	0·01	0·002	$0{\cdot}0^{4}8$
0·99	0·99	0·96	0·92	0·84	0·75	0·53	0·32	0·12	0·03	0·004

Table 6. *Lower bound in certain cases where repeated t confidence intervals are used only on differences* $\mu_i-\mu_j$ ($\nu = \infty$)

Conf. coeff. \ k	2	3	4	5	6	8	10	13	16	20
0·90	0·90	0·77	0·65	0·53	0·43	0·28	0·17	0·08	0·04	0·01
0·95	0·95	0·88	0·80	0·71	0·63	0·49	0·37	0·24	0·15	0·08
0·99	0·99	0·97	0·95	0·93	0·90	0·84	0·77	0·67	0·58	0·48

The sensitivity of the new method is exactly the same as that of the F-test of the hypothesis H in (5) at significance level α in the following sense: If the F-test accepts the hypothesis H that all the contrasts are zero, the new method will make statement (i) for all contrasts, that is, say no contrast is significantly different from zero; if the F-test rejects the hypothesis, the new method will make statements (ii) and (iii) for some contrasts, that is, say they are significantly different from zero (and positive or negative, respectively). This may readily be seen* from the geometrical picture introduced in §2.

To formulate the problem of the power of the method, we consider the probabilities of making each of the statements (i), (ii) and (iii). We might think at first that what we would like is to make statement (i) for a contrast whose true value is zero, (ii) for one whose true value is greater than some assigned θ_0, (iii) for one whose true value is less than $-\theta_0$, and that what we want are the probabilities of the desired statements in each case. This formulation is at once seen to be nonsensical, because for any contrast θ_1 whose true value is positive there exist constants h' and h'' such that the true value of $h'\theta_1$ is $<\theta_0$, the true value of $h''\theta_1$ is $>\theta_0$, and yet the same one of the statements (i), (ii) and (iii) will be made for all contrasts $h\theta_1$ with positive h. This difficulty can be avoided by laying our requirements on the suitably normalized contrasts; if we find what happens for the normalized contrasts, we will know what happens for all contrasts. The appropriate definition is that a contrast is normalized if the variance of its estimate is $C^2\sigma^2$, where C is a specified constant; for normalized contrasts and their estimates we then write ϑ and $\hat{\vartheta}$ instead of θ and $\hat{\theta}$.

To see how the above difficulty is met, suppose the structure of an experiment is such that the contrast $\mu_1-\mu_2$ is determined with greater precision than the contrast $\mu_4-\mu_5$. This is a property of the design which presumably was desired, and if we guarantee a certain probability P of detecting (that is, making statement (ii) for) a difference $\mu_1-\mu_2$ as great as θ_0, we should be satisfied with the same probability P of detecting a difference $\mu_4-\mu_5$ as great as a certain multiple of θ_0. The multiple assigned by the present method will later be seen to be the ratio of the standard errors of estimate of the two contrasts. Since this property may be expressed by saying that 'the method assures the same probability P of detecting all *normalized* contrasts as great as a certain bound ϑ_0', it indicates that we have a way of normalizing the contrasts which is suitable for this method.

While this motivates our normalizing the contrasts so that their estimates all have the same variance $C^2\sigma^2$, there is still the question of the choice of the constant C. Whether there is a satisfactory universal choice of C is dubious.† In any event, the question is of little practical importance, and we need not settle it in order to analyse the operating characteristic. If for the same experiment two statisticians choose C and C', then equivalent choices of the bounds ϑ_0 and ϑ_0' are related by $\vartheta_0' = (C'/C)\,\vartheta_0$.

A simple example illustrating these considerations will be helpful. Suppose we wish to design an experiment in a 'one-way lay-out' with $k = 6$, that we wish to take twice as many observations on the means μ_1, μ_2, μ_3 as on μ_4, μ_5, μ_6, and that we are primarily interested in the contrasts which are differences $\mu_i-\mu_j$ (for equal numbers of observations per mean we should then use Tukey's method). These contrasts will then be determined with three kinds of precision, namely, the precision of those like $\mu_1-\mu_2$, like $\mu_4-\mu_5$, and like $\mu_1-\mu_4$, the three standard errors of estimate being in the ratio $2^{\frac{1}{2}} : 4^{\frac{1}{2}} : 3^{\frac{1}{2}}$. Suppose we are satisfied with a probability of $P = 0{\cdot}90$ of detecting a difference $\mu_1-\mu_2$ as great as $1{\cdot}0\sigma$. We will then get the same probability $0{\cdot}90$ for a difference $\mu_4-\mu_5$ as great as $1{\cdot}0\sigma(4^{\frac{1}{2}}/2^{\frac{1}{2}}) = 1{\cdot}4\sigma$, or a difference $\mu_1-\mu_4$ as great as $1{\cdot}0\sigma(3^{\frac{1}{2}}/2^{\frac{1}{2}}) = 1{\cdot}2\sigma$. If for either of the last two differences this sensitivity is not considered sufficient then we should question whether we have chosen the correct design.

---

* In the notation of §2, H may be written $\eta = 0$. Thus the F-test accepts H if and only if the confidence sphere $\mathscr{S}$ covers the origin, and this is precisely the case where statement (i) will be made about all the contrasts.

† A possible universal choice of C might be the following: transform from $\hat{\mu}_1, \ldots, \hat{\mu}_k$ to $\hat{\xi}_1, \ldots, \hat{\xi}_k$ by an orthogonal transformation, such that $\hat{\xi}_k$ is the $\hat{\eta}_k$ of §2, and $\hat{\xi}_1, \ldots, \hat{\xi}_{k-1}$ are statistically independent. The set $\hat{\xi}_1, \ldots, \hat{\xi}_{k-1}$ then spans the space of the estimated contrasts. Let $\sigma_i^2 = \mathrm{var}\,(\hat{\xi}_i)$. Define C so that $C\sigma$ is the geometric mean of $\sigma_1, \ldots, \sigma_{k-1}$. C is then a function of the a_{ij} in (6) alone. It may be shown that in the case of independent $\hat{\mu}_i$ with equal variances σ^2/n this gives $C = n^{-\frac{1}{2}}$; more generally if the $\hat{\mu}_i$ are independent with respective variances σ^2/n_i this gives $C^{2(k-1)} = \overline{n}\Big/\prod_1^k n_i$, where $\overline{n} = \sum_1^k n_i/k$. In the example below, this leads to $C^2 = (\frac{3}{16})^{\frac{1}{5}}/n$, which does not particularly recommend itself from the practical computational point of view.

Let n be the number of observations on μ_4, μ_5, μ_6, and $2n$ the number on μ_1, μ_2, μ_3. While the choice of the normalizing constant C does not matter much and might even be left indefinite, we shall suppose for concreteness in this example that the contrasts are normalized so that their estimates all have the same variance as that of $\mu_1-\mu_2$, namely, σ^2/n. Since for any contrast

$$\operatorname{var}\left(\sum_1^k c_i\hat{\mu}_i\right) = \left(\frac{1}{2}\sum_1^3 c_i^2+\sum_4^6 c_i^2\right)\sigma^2/n,$$

the coefficients in a normalized contrast $\vartheta = \sum\limits_1^k c_i\mu_i$ must then satisfy the condition

$$\frac{1}{2}\sum_1^3 c_i^2+\sum_4^6 c_i^2 = 1.$$

Thus the normalized form of the contrast $\mu_4-\mu_5$ is $2^{-\frac{1}{2}}(\mu_4-\mu_5)$, and of $\mu_1-\mu_4$, $(\frac{2}{3})^{\frac{1}{2}}(\mu_1-\mu_4)$. Below we shall find how to choose n so that the probability is $0{\cdot}90$ of detecting a difference of $\mu_1-\mu_2$ equal to a given multiple of σ, say $1{\cdot}0\sigma = \vartheta_0$. The probability will then be the same for the value ϑ_0 of the contrast $2^{-\frac{1}{2}}(\mu_4-\mu_5)$ or the value ϑ_0 of the contrast $(\frac{2}{3})^{\frac{1}{2}}(\mu_1-\mu_4)$. This gives the statements in the last paragraph about the differences $\mu_4-\mu_5$ and $\mu_1-\mu_4$. In such an experiment the contrast consisting of the difference of the average of μ_1, μ_2, μ_3 and the average of μ_4, μ_5, μ_6 might also be of interest. Its normalized form is

$$(\tfrac{2}{9})^{\frac{1}{2}}\left(\sum_1^3 \mu_i-\sum_4^6 \mu_i\right),$$

and we find in a similar way that a difference of the averages $\sum\limits_1^3 \mu_i/3$ and $\sum\limits_4^6 \mu_i/3$ equal to $(\frac{1}{3})(\frac{9}{2})^{\frac{1}{2}}\vartheta_0 = 0{\cdot}7\sigma$ would also be detected with probability $0{\cdot}90$.

In the following calculations we shall symbolize by $t'(f;\,\delta)$ a non-central t variable with f D.F. and non-centrality parameter δ, that is, a variable distributed as the quotient of $z+\delta$ by $f^{-\frac{1}{2}}\chi$, where z is a standard normal deviate and χ is an independent χ variable with f D.F.

Statements (i), (ii) or (iii) are made for a contrast θ according as $\hat{\theta}/\hat{\sigma}_{\hat{\theta}}$ falls in the intervals $(-S, S)$, (S, ∞), or $(-\infty, -S)$. But the variable $\hat{\theta}/\hat{\sigma}_{\hat{\theta}}$ has the non-central t-distribution with ν D.F. and parameter $\delta = \theta/\sigma_{\hat{\theta}}$, where $\sigma_{\hat{\theta}}$ is given by (6). It follows that the probability of each of these statements being made depends only on the true value of $\theta/\sigma_{\hat{\theta}}$ for the particular contrast for which the statement is made, and on the constant S.

If the true value of θ is zero, the probability that we make the desired statement (i) is

$$\Pr\{-S \leqslant t'(\nu;\,0) \leqslant S\},$$

where $t'(\nu;\,0)$ denotes a central t variable with ν D.F. This may also be written

$$\Pr\{F(1,\nu) \leqslant (k-1)\,F_\alpha(k-1,\nu)\},$$

where $F(1, \nu)$ is an F variable with 1 and ν D.F. That this is at least $1-\alpha$ is evident from the confidence statement associated with (8). Indeed, as k increases from 2, this probability increases rapidly from $1-\alpha$ as indicated in Table 7, where 1 minus this probability is tabled for the case $\nu = \infty$.

Table 7. *Probability of not making statement* (i) *about a contrast whose true value is zero* ($\nu = \infty$)

k / α	2	3	4	5	6	8	10	13	16	20
0·10	0·10	0·032	0·012	0·0053	0·0024	$0{\cdot}0^{3}53$	$0{\cdot}0^{3}13$	$0{\cdot}0^{4}17$	$0{\cdot}0^{5}23$	$0{\cdot}0^{6}18$
0·05	0·050	0·014	0·0052	0·0021	$0{\cdot}0^{3}88$	$0{\cdot}0^{3}18$	$0{\cdot}0^{4}39$	$0{\cdot}0^{5}45$	$0{\cdot}0^{6}57$	$0{\cdot}0^{7}40$
0·01	0·010	0·0024	$0{\cdot}0^{3}76$	$0{\cdot}0^{3}27$	$0{\cdot}0^{3}10$	$0{\cdot}0^{4}17$	$0{\cdot}0^{5}32$	$0{\cdot}0^{6}31$	$0{\cdot}0^{7}32$	$0{\cdot}0^{8}18$

For any contrast θ the probability that we make statement (ii) is

$$\Pr\{t'(\nu;\, \theta/\sigma_{\hat\theta}) > S\}. \tag{18}$$

This is a strictly increasing function of the value of $\theta/\sigma_{\hat\theta}$. Suppose now the contrast is normalized; we then write it as ϑ and have $\sigma_{\hat\vartheta} = C\sigma$. Writing P for the probability (18) when $\vartheta = \vartheta_0$, we have

$$P = \Pr\{t'(\nu;\, \vartheta_0/(C\sigma)) > S\}. \tag{19}$$

From the tables of non-central t by Johnson & Welch (1940) we can find the value of $\delta = \vartheta_0/(C\sigma)$ for which P attains the values 0·01, 0·05, 0·1 (0·1) 0·9, 0·95, 0·99.

The probability of making statement (iii) for any contrast can of course be calculated as the probability of making statement (ii) for its negative.

In the example introduced above of the experiment with 6 means, $C = n^{-\frac{1}{2}}$. If we take $n = 10$ and use significance level $\alpha = 0{\cdot}05$, then $\nu = 84$, $S^2 = 5F_{0\cdot05}(5, 84) = (3{\cdot}41)^2$, and we find from (19) and the Johnson & Welch tables that for $P = 0{\cdot}90$, $\delta = 4{\cdot}72$, so $\vartheta_0 = 4{\cdot}72C\sigma = 1{\cdot}49\sigma$. This poor sensitivity is related to the extremely low risk (0·0009 in Table 7 for $\nu = \infty$) of making the other kind of error, namely, calling significantly different from zero the estimate of a contrast whose true value is zero. We supposed in this example that we desired the n that would give a $P = 0{\cdot}90$ for detecting a $\vartheta = 1{\cdot}0\sigma = \vartheta_0$. Since for large ν the parameter δ found from the tables does not vary much with ν we try for a first approximation for n the solution of $\vartheta_0/(C\sigma) = (1{\cdot}0)\, n^{\frac{1}{2}} = \delta$ with the former $\delta = 4{\cdot}72$. This gives $n = 22$. Calculation of ϑ_0 from the tables for $n = 21$, 22, as above for $n = 10$, shows that $n = 22$ is the required value. In the same way we find that if we desire $\vartheta_0 = 2{\cdot}0\sigma$ for $P = 0{\cdot}90$, a first approximation to the required n is $n = (4{\cdot}72/2{\cdot}0)^2 \doteqdot 6$; use of the tables also gives $n = 6$. In all these calculations we took $\alpha = 0{\cdot}05$.

Although the behaviour of the operating characteristic which we have now determined for any single contrast considered alone is of interest, it seems to the writer that in this method designed for judging all the contrasts it is of greater interest to determine the following two properties of the operating characteristic: *First, the probability* P_1 *of the event* $\mathscr{E}_1$ *that the desired statement* (i) *would be made for all the contrasts whose true values are zero, and secondly, the probability* P_2 *of the event* $\mathscr{E}_2$ *that the desired statement* (ii) *would be made for all normalized contrasts whose values are* $\geqslant \vartheta_0$ (*in which case the desired statement* (iii) *would also be made for all normalized contrasts whose values are* $\leqslant -\vartheta_0$).

We shall state here the results for the probabilities P_1 and P_2 and defer the proofs to the next section. It is obvious from (8) that if the hypothesis H in (5) is true, $P_1 = 1-\alpha$. If H is false

$$P_1 = \Pr\{F(k-2, \nu) \leqslant (k-1)(k-2)^{-1} F_\alpha(k-1, \nu)\}, \tag{20}$$

where $F(m, \nu)$ denotes an F variable with m and ν D.F., and $F_\alpha(m, \nu)$, its upper α point. It is interesting to note that P_1 depends only on whether H is true or false and not further on any unknown parameters. It is easy to see that if H is false $P_1 > 1-\alpha$. Table 8 gives some values of $1-P_1$; we remark that for $k = 2$ there are no zero contrasts other than $0 \cdot \mu_1 + 0 \cdot \mu_2$ if H is false, and so $1-P_1 = 0$ if H is false, trivially. We see now that this analogue of the Neyman-Pearson risk of a 'type I' error, namely, the probability $1-P_1$ of failing to make statement (i) for *all* zero contrasts, is the one the method controls in a satisfactory way, rather than the marginal probability for a *single* zero contrast considered above, which assumes the microscopic values indicated by Table 7.

Table 8. *Values of* $1-P_1$ *if* H *is false*

α	ν \ k	3	4	5	6	8	10	13	16	20
0·10	5	0·040	0·056	0·065	0·070	0·077	0·082	0·086	0·088	0·091
	7	0·038	0·053	0·061	0·067	0·074	0·079	0·083	0·086	0·089
	10	0·036	0·050	0·058	0·064	0·071	0·076	0·081	0·084	0·087
	20	0·034	0·047	0·055	0·060	0·067	0·072	0·077	0·080	0·083
	40	0·033	0·046	0·053	0·058	0·065	0·069	0·074	0·077	0·080
	∞	0·032	0·044	0·051	0·055	0·062	0·066	0·070	0·073	0·075
0·05	5	0·019	0·027	0·031	0·034	0·038	0·040	0·042	0·044	0·045
	7	0·018	0·025	0·029	0·032	0·036	0·039	0·041	0·042	0·044
	10	0·017	0·024	0·028	0·031	0·035	0·037	0·040	0·041	0·043
	20	0·016	0·022	0·026	0·028	0·032	0·035	0·037	0·039	0·041
	40	0·015	0·021	0·025	0·027	0·031	0·033	0·035	0·037	0·039
	∞	0·014	0·020	0·023	0·026	0·029	0·031	0·033	0·035	0·036
0·01	5	0·0036	0·0051	0·0060	0·0067	0·0074	0·0079	0·0084	0·0087	0·0089
	7	0·0033	0·0047	0·0056	0·0062	0·0070	0·0075	0·0080	0·0083	0·0086
	10	0·0030	0·0044	0·0052	0·0058	0·0066	0·0071	0·0076	0·0080	0·0083
	20	0·0027	0·0039	0·0047	0·0052	0·0060	0·0065	0·0070	0·0074	0·0078
	40	0·0026	0·0037	0·0044	0·0049	0·0056	0·0061	0·0066	0·0070	0·0074
	∞	0·0024	0·0034	0·0041	0·0045	0·0052	0·0056	0·0061	0·0063	0·0067

Let $\vartheta_{\max.}$ denote the largest of the true values of all the normalized contrasts. If $\vartheta_0 > \vartheta_{\max.}$ the probability P_2 is trivially 1. If $\vartheta_0 \leqslant \vartheta_{\max.}$ define the angle γ from

$$\sin\gamma = \vartheta_0/\vartheta_{\max.}. \tag{21}$$

Then it will be shown that in this case P_2 is the probability that

$$(\cot\gamma)\,\chi_1 + (S\nu^{-\frac{1}{2}}\operatorname{cosec}\gamma)\,\chi_2 \leqslant z_1 + (C\sigma)^{-1}\vartheta_0 \operatorname{cosec}\gamma, \tag{22}$$

where χ_1, χ_2, z_1 are statistically independent, χ_1 and χ_2 are χ variables with $k-2$ and ν D.F., respectively, and z_1 is a standard normal deviate. The probability P_2 does not seem exactly calculable from any existing tables, but an excellent approximation* may be obtained for moderate or large values of ν by replacing $\nu^{-\frac{1}{2}}\chi_2$ by a normal variable z_2 with mean 1 and variance $1/(2\nu)$. A similar approximation would not work so well for χ_1, since its number $k-2$ of D.F. might be quite small. This way we find P_2 approximately equal to the probability that

---

* Johnson & Welch (1940) used a similar method to approximate non-central t.

$$z+\delta \geqslant t_0(k-2)^{-\frac{1}{2}}\chi_1,$$

where
$$z = [1+(2\nu)^{-1}S^2\operatorname{cosec}^2\gamma]^{-\frac{1}{2}}(z_1 - z_2 S\operatorname{cosec}\gamma + S\operatorname{cosec}\gamma)$$

is a standard normal deviate and independent of χ_1;

$$\delta = [1+(2\nu)^{-1}S^2\operatorname{cosec}^2\gamma]^{-\frac{1}{2}}[(C\sigma)^{-1}\vartheta_0\operatorname{cosec}\gamma - S\operatorname{cosec}\gamma], \tag{23}$$

$$t_0 = [1+(2\nu)^{-1}S^2\operatorname{cosec}^2\gamma]^{-\frac{1}{2}}(k-2)^{\frac{1}{2}}\cot\gamma. \tag{24}$$

To this approximation P_2 can thus be expressed in terms of the non-central t-distribution as

$$\Pr\{t'(k-2;\ \delta) > t_0\}, \tag{25}$$

where δ and t_0 are given by (23) and (24). In the case where σ is known we can put $\nu = \infty$ in (22), so that the left member becomes $(\cot\gamma)\chi_1 + S\operatorname{cosec}\gamma$; we may also put $\nu = \infty$ in (23) and (24), and the above approximation for P_2 then becomes exact.

The exact probability P_2 and its approximation (25) depend on the parameter $\vartheta_{\text{max.}}$, which enters (22) through γ. As indicated in the example of the six means, we may be willing to specify ϑ_0 beforehand as a multiple of σ, say

$$\vartheta_0 = A\sigma, \tag{26}$$

where A is a given constant, but it is unpleasant to discover then that P_2 still depends on the unknown parameter

$$\psi = \vartheta_{\text{max.}}/\sigma. \tag{27}$$

The dependence of the exact probability P_2 of (22) on ψ may be made clear by writing (22) as

$$(\psi^2A^{-2}-1)^{\frac{1}{2}}\chi_1 + (\nu^{-\frac{1}{2}}S\psi/A)\chi_2 \leqslant z_1 + \psi C^{-1}, \tag{28}$$

where the variables χ_1, χ_2, z_1 have the distribution stated below (22). For $\psi > A$, P_2 is the probability of (28); for $0 \leqslant \psi \leqslant A$, P_2 trivially equals 1. This problem did not arise when we obtained in (19) the corresponding marginal probability for a single contrast alone whose value $= \vartheta_0$ (the non-centrality parameter δ there may be written A/C), as it does now when we consider the overall probability for all contrasts whose value is $\geqslant \vartheta_0$ (we could write $= \vartheta_0$ here also).

We suggest meeting this problem by using a lower bound P_2' for P_2. We shall prove that $P_2 = P_2(\psi)$ is monotone in ψ, decreasing from the value 1 for $\psi \leqslant A$ to a limiting value $P_2(\infty)$ as ψ increases from 0 to ∞. This means that if we are willing to assume ψ does not exceed a known ψ_1, we may use as the bound P_2' the value $P_2(\psi_1)$, which may be accurately approximated by (23), (24), (25) with $\gamma = \arcsin(A/\psi_1)$, and if we are unwilling to set a bound ψ_1 for ψ, or if we are satisfied with a somewhat larger but much more easily calculated bound, we may use $P_2' = P_2(\infty)$, whose value will now be stated.

The limiting value $P_2(\infty)$ is the probability that

$$\chi_1 + \nu^{-\frac{1}{2}}S\chi_2 \leqslant A/C, \tag{29}$$

where χ_1 and χ_2 are distributed as in (22). This may be accurately approximated by the same method that led to (25), with the result that $P_2(\infty)$ is approximately*

$$\Pr\{t'(k-2;\ \delta) > t_0\}, \tag{30}$$

---

* It is unfortunate in connexion with (30) and (25) that the Johnson & Welch tables (1940) do not go below 4 D.F., since 1, 2, 3 D.F. are needed for $k = 3, 4, 5$; it would be very desirable to have this extension of their tables.

where
$$\delta = (2\nu)^{\frac{1}{2}}\,[A(CS)^{-1}-1], \tag{31}$$
$$t_0 = (2\nu)^{\frac{1}{2}}\,(k-2)^{\frac{1}{2}}/S. \tag{32}$$

For very large values of ν we may use the following simple approximation to $P_2(\infty)$, obtained by replacing $\nu^{-\frac{1}{2}}\chi_2$ by 1 in (29):
$$\Pr\{\chi_1 \leqslant AC^{-1}-S\}.$$

To this approximation we may thus rapidly find for given β the A such that $P_2(\infty)$ attains the value $1-\beta$ for $\vartheta_0 = A\sigma$ by taking the square root of the upper β point from the χ^2 tables for $k-2$ D.F., say $\chi_{\beta;k-2}$, and computing
$$A \doteqdot C(S+\chi_{\beta;k-2}). \tag{33}$$

A still rougher approximation might be obtained by replacing S by its value for $\nu=\infty$, namely, $\chi_{\alpha;k-1}$, to give
$$A \doteqdot C(\chi_{\alpha;k-1}+\chi_{\beta;k-2}). \tag{34}$$

The approximation (34) becomes exact when σ is known; this is obtained by replacing $\hat{\sigma}$ by σ and ν by ∞ in the calculations.

To illustrate these results in the example we have carried along, we recall $k=6$, $C=n^{-\frac{1}{2}}$, $\alpha=0{\cdot}05$. Suppose we wish a bound P_2' for P_2 and decide to use $P_2' = P_2(\infty)$. If we take $n=10$, then $\nu=84$, and entering the Johnson & Welch tables (1940) with $t_0 = 7{\cdot}61$ from (32) we find that a $P_2(\infty) = 0{\cdot}90$ (to the approximation of (30)) is attained for $\delta = 10{\cdot}8$, and this gives $A = 1{\cdot}98$ from (31), that is, for $\vartheta_0 = 1{\cdot}98\sigma$. The approximation (33) much more quickly gives $A = 6{\cdot}20C = 6{\cdot}20n^{-\frac{1}{2}}$ or $A = 1{\cdot}96$. The approximation (34) gives $A = 1{\cdot}93$. To find the n which gives a $P_2(\infty) = 0{\cdot}90$ for $\vartheta_0 = 1{\cdot}0\sigma$, we first use the previous approximation from (33), $A = 6{\cdot}20n^{-\frac{1}{2}}$, which is calculated with an S based on the now wrong $\nu = 84$, but then S is not sensitive to changes in ν for large ν. This gives $n = 38$ for $A = 1{\cdot}0$; the more correct formula (30) also gives $n = 38$. For most applications we would probably have to compromise on a larger A to get a more feasible n. In a similar way for $n=5$, $\nu=39$, $\alpha = 0{\cdot}05$, $C = 5^{-\frac{1}{2}}$, we find first $S = 3{\cdot}50$, then $A = 2{\cdot}81$ from the approximation (33); the more accurate (30) gives $A = 2{\cdot}87$.

5. DERIVATION OF P_1, P_2, ETC.

The probabilities P_1 and P_2 can be found neatly by continuing the geometric approach of §2. Extending the notation there, denote by $|\eta|$ the length of the vector η, $|\eta|^2 = \sum_1^{k-1}\eta_i^2$. Since the value ϑ of a normalized contrast is the projection of η on d, it is evident that
$$|\eta| = \vartheta_{\max.} \tag{35}$$

Suppose for the moment $\eta \neq 0$ (equivalent to H false). If we imagine marking off the projection of η on d for each unit vector d to get a polar co-ordinate graph of the totality of values of the normalized contrasts, we see the graph consists of a sphere which has the vector η as a diameter. The zero-valued contrasts for which we desire to make statement (i) constitute the $(k-2)$-dimensional set of vectors d in the tangent plane $\mathscr{P}$ to the sphere at the origin. If $\eta = 0$ this picture collapses in an obvious way.

Consider now the totality of normalized contrasts whose values are $\geqslant \vartheta_0$, for which we wish to make the statement (ii). There are none if $|\eta| < \vartheta_0$. If $|\eta| \geqslant \vartheta_0$ they fill a circular cone $\mathscr{C}$ (by 'cone' we mean throughout *one nappe* of a cone) with axis along η, vertex at the origin, and elements making an angle $\frac{1}{2}\pi-\gamma$ with η, where γ is defined by (22). The cone $\mathscr{C}'$ obtained by reflecting $\mathscr{C}$ in the origin is filled by the normalized contrasts whose values are $\leqslant -\vartheta_0$.

In the same y-space the estimated values of the normalized contrasts have a similar graph, the diameter of the sphere being $\hat{\eta}$. From the interpretation of the confidence interval (13) for the value ϑ of any normalized contrast d as the projection of the sphere $\mathscr{S}$ on d, it is easy to see that if $|\hat{\eta}| < SC\hat{\sigma}$ the statement (i) is made for all the contrasts, while if $|\hat{\eta}| \geqslant SC\hat{\sigma}$ the normalized contrasts for which we make the statement (ii) fill a cone $\mathscr{D}$ with axis along $\hat{\eta}$, vertex at the origin, and elements making an angle arccos $(SC\hat{\sigma}/|\hat{\eta}|)$ with $\hat{\eta}$, and the normalized contrasts for which we make the statement (iii) fill the reflexion $\mathscr{D}'$ of $\mathscr{D}$ in the origin.

If $\eta \neq 0$, the event $\mathscr{E}_1$ of probability P_1 happens if and only if the projection of the sphere $\mathscr{S}$ on the plane $\mathscr{P}$ covers the origin. It is convenient now to rotate the axes in the y-space so that the vector η lies along the positive y_{k-1}-axis. Denote the new coordinates of $\hat{\eta}$ by $\hat{\zeta}_1, \ldots, \hat{\zeta}_{k-1}$, and of η by $\zeta_1, \ldots \zeta_{k-1}$, respectively. Then $\hat{\zeta}_1, \ldots, \hat{\zeta}_{k-1}, \hat{\sigma}^2$ will be independent, the $\hat{\zeta}_1$ will be normal with variance $C^2\sigma^2$, the means $\zeta_1, \ldots, \zeta_{k-2}$ will all be zero, while $\zeta_{k-1} = |\eta|$. The plane $\mathscr{P}$ now has the equation $y_{k-1} = 0$. The distance of the centre of $\mathscr{S}$ from the y_{k-1}-axis is $\left(\sum_1^{k-2} \hat{\zeta}_i^2\right)^{\frac{1}{2}}$, and so the projection of $\mathscr{S}$ on $\mathscr{P}$ will cover the origin if and only if this distance does not exceed the radius of $\mathscr{S}$,

$$\sum_1^{k-2} \hat{\zeta}_i^2 \leqslant S^2C^2\hat{\sigma}^2,$$

or

$$\frac{\sum_1^{k-2} \hat{\zeta}_i^2/(k-2)}{C^2\hat{\sigma}^2} \leqslant (k-1)(k-2)^{-1} F_\alpha(k-1, \nu).$$

Since the random variable on the left side of the latter inequality has the F-distribution with $k-2$ and ν D.F., we have now succeeded in deriving (20).

If $|\eta| \geqslant \vartheta_0$, the event $\mathscr{E}_2$ happens if and only if the fixed cone $\mathscr{C}$ lies entirely inside the random cone $\mathscr{D}$. It may be verified that this is equivalent to the confidence sphere $\mathscr{S}$ lying entirely within a cone $\mathscr{F}$ with axis along η, vertex at the origin, and elements making with η an angle γ defined by (21). The sphere $\mathscr{S}$ will touch the cone $\mathscr{F}$ on the inside if the distance of its centre from the y_{k-1}-axis plus sec γ times the radius of the sphere equals $\zeta_{k-1}\tan\gamma$. Thus the event $\mathscr{E}_2$ happens if and only if

$$\left(\sum_1^{k-2} \hat{\zeta}_i^2\right)^{\frac{1}{2}} + CS\hat{\sigma}\sec\gamma \leqslant \hat{\zeta}_{k-1}\tan\gamma.$$

If we now divide this inequality through by $C\sigma\tan\gamma$ we get the desired result for P_2 stated in connexion with (22).

The monotone behaviour of $P_2(\psi)$ stated in §4 was suggested by the above geometric picture of the sphere $\mathscr{S}$ and the cone $\mathscr{F}$, and may be established by geometric arguments. It will be simpler to give here an analytic proof utilizing the inequality (22). Denote by $f(\gamma)$ the conditional probability, given χ_2, that this inequality is satisfied. Clearly it will suffice to show that for all $\chi_2 > 0$, $f(\gamma)$ is a monotone decreasing function of ψ for $\psi > A$, or since cosec $\gamma = \psi/A$, that the derivative $f'(\gamma)$ is positive.

Let us write the case of equality in (22) as

$$\chi_1 = z_1\tan\gamma + B\sec\gamma, \tag{36}$$

where

$$B = AC^{-1} - \nu^{-\frac{1}{2}}S\chi_2.$$

Because of the statistical independence of χ_1, χ_2, z_1, the conditional probability $f(\gamma)$ may be found by treating χ_2 as a constant and working with the joint distribution of χ_1 and z_1 in the z_1, χ_1-plane: $f(\gamma)$ is the amount of probability in this plane above the z_1-axis and below the line (36). We shall drop the subscripts from χ_1 and z_1 in the rest of this calculation. The z-intercept of the line (36) is $-B \operatorname{cosec} \gamma$, and the probability $f(\gamma)$ may thus be expressed as the integral

$$f(\gamma) = \int_{-B \operatorname{cosec} \gamma}^{\infty} p_1(z)\, g(z, \gamma)\, dz,$$

where

$$g(z, \gamma) = \int_0^{z \tan \gamma + B \sec \gamma} p_2(\chi)\, d\chi,$$

and $p_1(z)$ and $p_2(\chi)$ are the densities of a standard normal deviate and of a χ variable with $k-2$ D.F., respectively, and do not depend on γ. Now

$$f'(\gamma) = -p_1(-B \operatorname{cosec} \gamma)\, g(-B \operatorname{cosec} \gamma, \gamma)\, \partial(-B \operatorname{cosec} \gamma)/\partial\gamma + \int_{-B \operatorname{cosec} \gamma}^{\infty} p_1(z)\, [\partial g(z, \gamma)/\partial\gamma]\, dz,$$

and

$$\begin{aligned} \partial g(z, \gamma)/\partial\gamma &= p_2(z \tan \gamma + B \sec \gamma)\, \partial(z \tan \gamma + B \sec \gamma)/\partial\gamma \\ &= (z \sec^2 \gamma + B \sec \gamma \tan \gamma)\, p_2(z \tan \gamma + B \sec \gamma). \end{aligned}$$

But $g(-B \operatorname{cosec} \gamma, \gamma) = 0$, and hence

$$f'(\gamma) = \int_{-B \operatorname{cosec} \gamma}^{\infty} (z \sec^2 \gamma + B \sec \gamma \tan \gamma)\, p_1(z)\, p_2(z \tan \gamma + B \sec \gamma)\, dz. \tag{37}$$

If $B \leqslant 0$ the integrand is $\geqslant 0$ on the range of integration and then clearly $f'(\gamma) > 0$. If $B > 0$ we transform the integral (37) by the substitution $w = z \sec \gamma + B \tan \gamma$ to get

$$f'(\gamma) = \int_{-B \cot \gamma}^{\infty} p_1(w \cos \gamma - B \sin \gamma)\, p_2(w \sin \gamma + B \cos \gamma)\, dw.$$

By utilizing the explicit forms of the densities p_1 and p_2 this may be written

$$f'(\gamma) = \int_{-B \cot \gamma}^{\infty} v(w)\, dw, \tag{38}$$

where

$$v(w) = D[u(w)]^{k-3}\, w\, e^{-\frac{1}{2}w^2},$$

D is a positive constant, and

$$u(w) = w \sin \gamma + B \cos \gamma \tag{39}$$

is positive for $z > -B \cot \gamma$. If we break the range of integration in (38) up into the intervals $(-B \cot \gamma, 0)$, $(0, B \cot \gamma)$, $(B \cot \gamma, \infty)$, and drop the last, where $v(w) > 0$, we get

$$f'(\gamma) > \int_0^{B \cot \gamma} [v(-w) + v(w)]\, dw. \tag{40}$$

For $0 < w < B \cot \gamma$, $u(-w)$ is seen from (39) to be $< u(w)$, hence $|v(-w)| \leqslant v(w)$ since $k \geqslant 3$, and so the integrand in (40) is $\geqslant 0$ and therefore $f'(\gamma) > 0$.

To calculate the limiting value $P_2(\infty)$ we revert to the geometric picture. To make $\psi \to \infty$ we may let $|\eta| \to \infty$ with fixed σ. For $|\eta| > \vartheta_0$, we recall that $P_2(\psi)$ is the probability that the sphere $\mathscr{S}$ lies inside the cone $\mathscr{F}$. The cone $\mathscr{F}$ may be constructed by circumscribing a cone with vertex at the origin about a sphere $\mathscr{T}$ with centre at $|\eta|$ on the y_{k-1}-axis and radius ϑ_0. As $|\eta| \to \infty$ this picture becomes indeterminate. However, we clearly get the same result if we hold the sphere $\mathscr{T}$ and the joint distribution of $\hat{\zeta}_1, \ldots, \hat{\zeta}_{k-1}, \hat{\sigma}^2$ fixed and let the vertex of the circumscribed cone $\mathscr{F}$ go to $-\infty$ on the y_{k-1}-axis. The limiting figure for the cone $\mathscr{F}$ is now the cylinder circumscribed about $\mathscr{T}$ with elements parallel to the y_{k-1}-axis. The probability $P_2(\infty)$ that the sphere $\mathscr{S}$ fall in this cylinder is the probability that

$$\left(\sum_{1}^{k-2}\hat{\zeta}_i^2\right)^{\frac{1}{2}} + CS\hat{\sigma} \leqslant \vartheta_0.$$

Division of this inequality by $C\sigma$ yields the desired (29) and concludes the derivations.

The writer is indebted to Mr Seiji Sugihara and Mr Judah Rosenblatt for making the numerical calculations. Tables of Merrington & Thompson (1943) were used in the calculation of Tables 1, 2, 4 and 8, tables of Pearson & Hartley (1942) for Tables 1 and 6, tables of May (1952) for Table 2, tables of Merrington (1942) for Table 4, tables of Hartley & Pearson (1950) for Table 5, tables of Thompson (1941) and the W.P.A. tables (1942) for Table 7, and tables of Karl Pearson (1934) for Table 8.

REFERENCES

EISENHART, C. (1947). The assumptions underlying the analysis of variance. *Biometrics*, **3**, 1–21.

FISHER, R. A. (1935). *The Design of Experiments*, §24. Edinburgh: Oliver and Boyd.

FISHER, R. A. & YATES, F. (1943). *Statistical Tables*, Table 23. Edinburgh: Oliver and Boyd.

HARTLEY, H. O. & PEARSON, E. S. (1950). Tables of the χ^2-integral and of the cumulative Poisson distribution. *Biometrika*, **37**, 313–25.

JOHNSON, N. L. & WELCH, B. L. (1940). Applications of the noncentral t-distribution. *Biometrika*, **31**, 362–89.

MAY, JOYCE M. (1952). Extended and corrected tables of the upper percentage points of the 'Studentized' range. *Biometrika*, **39**, 192–3.

MERRINGTON, M. (1942). Table of percentage points of the t-distribution. *Biometrika*, **32**, 300.

MERRINGTON, M. & THOMPSON, C. M. (1943). Tables of the percentage points of the inverted beta (F) distribution. *Biometrika*, **33**, 73–88.

MOOD, A. M. (1950). *Introduction to the Theory of Statistics*, §§14·9–14·11. New York: McGraw-Hill.

NANDI, H. K. (1951). On the analysis of variance test. *Bull. Calcutta Statist. Ass.* **3**, 103–14.

NEWMAN, D. (1939). The distribution of range in samples from a normal population, expressed in terms of an independent estimate of standard deviation. *Biometrika*, **31**, 20–30.

NEYMAN, J. & PEARSON, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans.* A, **231**, 289–337.

PEARSON, E. S. & HARTLEY, H. O. (1942). The probability integral of the range in samples of n observations from a normal population. *Biometrika*, **32**, 301–10.

PEARSON, K. (1934). *Tables of the Incomplete Beta-Function*. London: Biometrika Office.

SCHEFFÉ, H. (1952). An analysis of variance for paired comparisons. *J. Amer. Statist. Ass.* **47**, 381–400.

THOMPSON, C. M. (1941). Table of percentage points of the χ^2-distribution. *Biometrika*, **32**, 187–91.

TUKEY, J. W. (1951). Quick and dirty methods in statistics. Part II. Simple analyses for standard designs. *Proc. Fifth Annual Convention, Amer. Soc. for Quality Control*, pp. 189–97.

W.P.A. TABLES (1942). *Tables of Probability Functions*, vol. **2**. Washington, D.C.: Nat. Bur. Standards.

THE ESTIMATION AND COMPARISON OF STRENGTHS OF ASSOCIATION IN CONTINGENCY TABLES

By A. STUART

Division of Research Techniques, London School of Economics

1. Introduction

In testing the significance of an observed association between two characteristics in a contingency table, χ^2 is generally used. When, however, we wish to *measure* strength of association, some text-books recommend the use of Karl Pearson's coefficient of contingency, a function of χ^2 which may be referred to the conventional range $(-1, +1)$, although it cannot attain the limits of this range. Kendall (1948*a*, chapter 13) gives references to the work on the coefficient, largely by Karl Pearson himself, and points out the difficulty of obtaining its sampling variance on the hypothesis of independence.

Yates (1948) proposed a test based on scores for association of characteristics of the kind considered in this paper, but its distribution, as he points out, is only obtained on the hypothesis of independence, and it becomes progressively more inaccurate as the degree of association increases. Williams (1952) has recently considered various tests based on scores.

In this paper, a coefficient is proposed which is independent of scoring systems. It is, in fact, a suitably modified form of Kendall's rank correlation coefficient, which depends only on ordinal properties. It is shown how the existing theory of the coefficient may be used to estimate the population association, to set confidence limits for it, and also to test the difference in the coefficients calculated for two contingency tables.

Since the proposed coefficient depends on rank order, a condition for its applicability is that the rows and columns of the contingency table fall into a natural order. This is, in fact, generally the case. From one point of view the proposed coefficient may therefore be regarded as availing itself of more of the information supplied by the contingency table than does the contingency coefficient, which is invariant under interchanges of rows or columns.

The confidence interval and the significance test obtained by use of the proposed coefficient are conservative, in the sense that a lower bound is provided for the confidence coefficient, and an upper bound for the probability of wrongly rejecting the hypothesis tested. This is a consequence of the fact that only an upper bound is known for the sampling variance of the coefficient, and this is in general a poor upper bound, although sometimes attainable. However, a conservative test is better than no test at all, and this test should be useful in many practical situations, especially where sample sizes are large.

2. Contingency tables and rankings

Kendall (1949) has suggested that his rank correlation coefficient t may be calculated for a contingency table whose rows and columns are ordered by the criteria of classification. An $r \times s$ table with a grand total of n is regarded as two rankings of n objects according to characteristics for one of which only r separate ranks are distinguished, and for the other only s separate ranks. Looked at in this way, the marginal totals are simply the numbers of objects tied at each level. There is thus a one-one correspondence between any contingency table and a pair of rankings.

The original context of this suggestion was the estimation of the product-moment correlation parameter in the case of non-normal variation, but there is no reason why the coefficient t should not be used in its own right as a measure of association for contingency tables in which there is a natural order for rows and columns.

3. THE DENOMINATOR OF t

The test commonly used to test the significance of an observed value of t, where there are no ties in the rankings, is described by Kendall (1948*b*). Like the χ^2 test for a contingency table, it is made with reference to a hypothetical population obtained by the permutation of sample observations. In calculating the coefficient, we may therefore use Daniels's (1944) definition of t in the universe of sample permutations. If r_i, r_j are the ranks allotted to the ith and jth objects ($i<j$), and a_{ij}, b_{ij} are scores, one defined on each ranking, given by

$$a_{ij}, b_{ij} = \begin{cases} +1 & (r_i < r_j), \\ -1 & (r_i > r_j), \end{cases}$$

this definition is

$$t = \frac{\Sigma a_{ij} b_{ij}}{(\Sigma a_{ij}^2)^{\frac{1}{2}} (\Sigma b_{ij}^2)^{\frac{1}{2}}}. \tag{1}$$

The summations in (1) extend over all unequal suffixes. The property of the coefficient

$$-1 \leqslant t \leqslant +1 \tag{2}$$

is seen from (1) to be a consequence of the Cauchy inequality, the equalities in (2) being attainable only when the members of each set of scores are all $+1$ or all -1. When there are no ties, (1) is equivalent to the more normal definition

$$t = \frac{2S}{n(n-1)}, \tag{3}$$

where S is a function of the number of inversions formed by the two rankings.

When we come to consider rankings containing ties, we must further define

$$a_{ij}, b_{ij} = 0 \quad (r_i = r_j),$$

and alternative forms are now possible for t, for although it remains true that $\Sigma a_{ij} b_{ij} = 2S$, the denominators of (1) and (3) no longer agree. If we are interested simply in testing whether association exists, the discussion of which form of denominator to use for t is unnecessary; the test may be carried out directly on the numerator. But if we are concerned to *measure* association, it is of some importance to choose a denominator so that the limits of the conventional range $(-1, +1)$ are at least sometimes attainable. With this in mind, we may now examine the possibilities.

(*a*) If, when there are ties present, we continue to use $n(n-1)$ as the denominator, as in (3), we obtain the coefficient denoted by Kendall (1948*b*) as t_a.

(*b*) If, on the other hand, we continue to use (1), the coefficient is denoted by t_b. The fact that some of the a_{ij}, b_{ij} are zero implies that the denominator is decreased in the tied, as compared to the untied, case. Nevertheless, it follows from the Cauchy inequality that the equalities in (2) are not attainable by t_b except in the trivial case when all the scores are zero. It further follows that since t_a has the same numerator but a larger denominator than t_b, and t_b cannot attain ±1, t_a certainly cannot attain ±1. (This is analogous to the similar deficiency of the contingency coefficient.) Where the number of ties is large, the deficiency will be serious. Since we are particularly interested in contingency tables, which often have large marginal totals, we must consider whether a more appropriate form of denominator exists.

(*c*) Consider the maximum (positive or negative) numerator $\Sigma a_{ij} b_{ij}$ which can be produced by a sample of n observations arranged in a $r \times s$ table. This will be attained when all observations lie in cells in a longest diagonal of the table, and the frequencies in these cells are equal, or as nearly equal as possible. Any move away from this position will decrease the total score $\Sigma a_{ij} b_{ij}$. A longest diagonal contains m cells, where m is the lesser of r and s.

If n is a multiple of m, the maximum sample score is

$$\Sigma a_{ij} b_{ij} = 2\left(\frac{n}{m}\right)^2 \{1+2+\ldots+(m-1)\} = n^2\left(\frac{m-1}{m}\right). \tag{4}$$

(If $n = m$, we are back to the untied case, and (4) reduces to $n(n-1)$, as it should.) Thus (4) is an attainable upper bound for $\Sigma a_{ij} b_{ij}$.

When n is not a multiple of m, (4) remains an upper bound, but cannot be attained. This is a quite unimportant deficiency in practice, when n is often large and m very small, and the residue of n to modulus m necessarily small.

We therefore define

$$t_c = \frac{\Sigma a_{ij} b_{ij}}{n^2 \dfrac{(m-1)}{m}} = \frac{2S}{n^2 \dfrac{(m-1)}{m}}, \tag{5}$$

whence

$$t_c = \frac{(n-1)}{n} \frac{m}{(m-1)} t_a. \tag{6}$$

It follows that t_c can sometimes attain, and for large n can generally almost attain, ± 1.

4. THE SAMPLING PROPERTIES OF t_c

Remembering the correspondence between a pair of rankings and a contingency table, let us consider the cell frequencies of an $r \times s$ table,

$$f_{ik} \quad (i = 1, \ldots, r;\ k = 1, \ldots, s),$$

as having been sampled from a population $r \times s$ table with frequencies F_{ik}. Then

$$t_a = \frac{1}{n(n-1)} \Sigma a_{ij} b_{kl} f_{ik} f_{jl}$$

and

$$E(t_a) = \frac{1}{n(n-1)} \Sigma a_{ij} b_{kl} E(f_{ik} f_{jl}),$$

which, by the ordinary formula for sampling from a finite multinomial population, becomes

$$= \frac{1}{n(n-1)} \Sigma a_{ij} b_{kl} \frac{n(n-1)}{N(N-1)} F_{ik} F_{jl}$$

$$= \frac{1}{N(N-1)} \Sigma a_{ij} b_{kl} F_{ik} F_{jl} = \tau_a, \tag{7}$$

where τ_a is defined for the population exactly as is t_a for the sample. Thus t_a is unbiased for τ_a.

Now Hoeffding (1948) has shown that for large N, the sampling distribution of t_a tends to the normal form as n increases, and its sampling variance obeys the inequality, first given by Daniels & Kendall (1947) for the untied case,

$$\operatorname{var} t_a \leqslant \frac{2}{n}(1-\tau_a^2). \tag{8}$$

If we now define, analogously to (6),

$$\tau_c = \frac{(N-1)}{N}\frac{m}{(m-1)}\tau_a,$$

and use (6), (7) and (8), we find

$$E(t_c) = \frac{(n-1)}{n}\frac{N}{(N-1)}\tau_c,$$

$$\operatorname{var} t_c \leqslant \frac{2}{n}\left\{\left[\frac{(n-1)}{n}\frac{m}{(m-1)}\right]^2 - \left[\frac{(n-1)}{n}\frac{N}{(N-1)}\tau_c\right]^2\right\},$$

which for large N and n reduce to

$$\left.\begin{aligned} E(t_c) &= \tau_c, \\ \operatorname{var} t_c &\leqslant \frac{2}{n}\left\{\left(\frac{m}{m-1}\right)^2 - \tau_c\right\}^2. \end{aligned}\right\} \tag{9}$$

As usual, we may estimate τ_c in the sampling variance by t_c.

The net effect of using t_c is therefore to multiply the value of our measure of association by $m/(m-1)$ approximately, and increase its sampling variance appropriately, while its property of unbiased estimation of its population correspondent τ becomes asymptotic.

(9) may be used to estimate τ_c, and to set conservative confidence limits for it. Similarly, a conservative test for the significance of the difference between two observed values of t_c follows immediately. Conservative procedures are forced upon us by the fact that only an upper bound to the sampling variance of the coefficient can be obtained. Daniels (1950) has shown that this upper bound, though sometimes attainable, is in general a poor one. Nevertheless, when n is at all large, even these conservative limits are close enough together for many practical purposes.

5. The calculation of S

Arrange the columns of the table so that they are ordered from left to right, and the rows so that they are ordered in the same sense from top to bottom. That is to say, our 'origin' is at the top left corner of the table.

For each cell in the table, multiply its frequency:

(i) positively, by the frequencies of all cells lying below it and to its right;

(ii) negatively, by the frequencies of all cells lying below it and to its left;

ignore all cells lying above it, or in the same row, or in the same column below it.

S is the sum obtained after applying this process to every cell in the table. Since the cells in the last row have none below them, their score is always zero.

6. Example

Tables 1 and 2, based on case records of the eye-testing of employees in Royal Ordnance factories in 1943–6, have been constructed from data very kindly made available by the Association of Optical Practitioners.

First, we calculate the t_c association coefficient for each table. Using the method described in §5, we find that for men (Table 1), $S = +2{,}480{,}223$, so that

$$t_c = \frac{2S}{n^2\left(\frac{m-1}{m}\right)} = \frac{2\times 2{,}480{,}223}{(3242)^2\times\frac{3}{4}} = +0{\cdot}629.$$

For women (Table 2), similarly, $S = +13{,}264{,}256$, so that

$$t_c = \frac{2\times 13{,}264{,}256}{(7477)^2\times\frac{3}{4}} = +0{\cdot}633.$$

Table 1. 3242 *men aged* 30–39: *unaided distance vision*

Right eye \ Left eye	Highest grade	Second grade	Third grade	Lowest grade	Total
Highest grade	821	112	85	35	1053
Second grade	116	494	145	27	782
Third grade	72	151	583	87	893
Lowest grade	43	34	106	331	514
Total	1052	791	919	480	3242

Table 2. 7477 *women aged* 30–39: *unaided distance vision*

Right eye \ Left eye	Highest grade	Second grade	Third grade	Lowest grade	Total
Highest grade	1520	266	124	66	1976
Second grade	234	1512	432	78	2256
Third grade	117	362	1772	205	2456
Lowest grade	36	82	179	492	789
Total	1907	2222	2507	841	7477

These values are very close together, and in the ordinary course of events we would not, perhaps, bother to test their difference. For illustrative purposes, however, we may carry through the calculations.

Our estimate of the upper bound to the sampling variance of t_c for men is, by (9),

$$\frac{2}{3242}\{(\tfrac{4}{3})^2-(0{\cdot}629)^2\} = 0{\cdot}000853,$$

and for women, similarly, it is

$$\frac{2}{7477}\{(\tfrac{4}{3})^2-(0{\cdot}633)^2\} = 0{\cdot}000368.$$

The values of the coefficient, and their difference, with the estimates of the appropriate maximum standard errors, and the conservative confidence limits based on two standard errors, are therefore as follows:

	t_c	Maximum standard error	2 S.E. conservative confidence limits
Men	+0·629	0·029	0·571 to 0·687
Women	+0·633	0·019	0·595 to 0·671
Difference	0·004	0·035	—

It follows, as we anticipated, that the difference in strength of association has not been shown to be significant.

REFERENCES

Daniels, H. E. (1944). *Biometrika*, **33**, 129.
Daniels, H. E. & Kendall, M. G. (1947). *Biometrika*, **34**, 197.
Daniels, H. E. (1950). *J.R. Statist. Soc.* B, **12**, 171.
Hoeffding, W. (1948). *Ann. Math. Statist.* **19**, 293.
Kendall, M. G. (1948*a*). *The Advanced Theory of Statistics*, vol. **1**, 4th ed. London: Charles Griffin.
Kendall, M. G. (1948*b*). *Rank Correlation Methods*. London: Charles Griffin.
Kendall, M. G. (1949). *Biometrika*, **36**, 177.
Williams, E. J. (1952). *Biometrika*, **39**, 274.
Yates, F. (1948). *Biometrika*, **35**, 176.

A SEQUENTIAL TEST FOR RANDOMNESS

BY P. G. MOORE
University College, London

1. The problem frequently arises of deciding whether a sequence of observations, each observation falling into one of two alternative categories or types, occurs in a random order. David (1947) has considered the 'group' test in some detail. The hypothesis, H_0, that we are considering here states that there is randomness within the sequence against an alternative, H_1, that there is dependence of the kind found in a simple Markoff chain. David's test is based on the number of groups of a common type which the observations form when placed in order in the sequence.

In this paper we are going to deal with another form of this procedure which leads to a sequential test. This has obvious advantages in that the amount of data which must be collected in order to obtain a decision is not fixed beforehand. It may be that the first few values will suffice to make a decision one way or the other and only in cases which lie in between will a long sequence of observations be necessary. This form of procedure may well mean a large saving of time, money and materials.

2. In the sequence of alternatives we write 1 for the happening of a certain event and 0 for its negation. Let E_s represent the sth event, which may either be a 1 or a 0. The theoretical model envisaged is that of a simple Markoff chain. The general principles of such a chain have been discussed at some length by Fréchet (1938) and the notation that we will use is that

$$\begin{aligned} &\Pr\{E_1 = 1\} = P, && \Pr\{E_1 = 0\} = Q, \\ &\Pr\{E_s = 1 \mid E_{s-1} = 1\} = p_1, && \Pr\{E_s = 0 \mid E_{s-1} = 1\} = q_1, \\ &\Pr\{E_s = 1 \mid E_{s-1} = 0\} = p_2, && \Pr\{E_s = 0 \mid E_{s-1} = 0\} = q_2, \end{aligned}$$

where
$$P+Q = 1, \quad p_1+q_1 = 1, \quad p_2+q_2 = 1.$$

These equations define the general set-up. The hypothesis of randomness is a special case of this Markoff chain when the probabilities of the two possible results of a trial are the same whatever the result of the previous trial. This means that $p_1 = p_2 = p$ (say) and hence, of course, $q_1 = q_2$.

It is necessary to make some assumption concerning the start of the sequence. If we assume that the sequence's start is at a randomly selected point in a longer sequence which also obeys the same probability model as has been given, then

$$\Pr\{E_1 = 1\} = \Pr\{E_1 = 1, E_0 = 1\} + \Pr\{E_1 = 1, E_0 = 0\},$$

where E_1 is the event chosen to start the sequence and E_0 is the previous event. Hence

$$P = p_1 P + p_2 Q,$$

or
$$P(1-p_1) = p_2 Q \quad \text{giving} \quad p_2 = Pq_1/Q.$$

We will denote the hypothesis of randomness (specifying a particular value of p) by H_0 and a Markoff chain set-up as defined by the equations given earlier as H_1. Then it is known (David, 1947, p. 335) that, under the assumptions made above, if there be an even number of groups ($2t$), then

$$\Pr\{t \mid r_1, r_2, H_0\} = k\, {}^{r_1-1}C_{t-1}\, {}^{r_2-1}C_{t-1}, \tag{1}$$

$$\Pr\{t \mid r_1, r_2, H_1\} = k'\, {}^{r_1-1}C_{t-1}\, {}^{r_2-1}C_{t-1} \left(\frac{p_2 q_1}{p_1 q_2}\right)^t, \tag{2}$$

where r_1 and r_2 are the numbers of 1's and 0's respectively and k and k' are constants such that $\Sigma \Pr\{2t\} = 1$, where Σ denotes summation over all possible values of t.

For the case of an odd number of groups the expressions are more complicated. If there are $2t+1$ groups, then

$$\Pr\{t \mid r_1, r_2, H_0\} = l({}^{r_1-1}C_t\,{}^{r_2-1}C_{t-1} + {}^{r_1-1}C_{t-1}\,{}^{r_2-1}C_t)$$

$$\Pr\{t \mid r_1, r_2, H_1\} = l'\left(\frac{p_2q_1}{p_1q_2}\right)^t \left[\frac{P}{p_1}\,{}^{r_1-1}C_t\,{}^{r_2-1}C_{t-1} + \frac{Q}{q_2}\,{}^{r_1-1}C_{t-1}\,{}^{r_2-1}C_t\right],$$

l and l' being constants such that the sum of the probabilities over all possible values of t is unity. Moore (1949) has shown that for t large the distributions of $\Pr\{t\}$ under either H_0 or H_1 approximate to one another. That is to say, that whether there be an odd or even number of groups, the distribution of $\Pr\{t \mid H\}$ is approximately the same. Hence for the purposes of this paper attention will be focused on the case of an even number of groups, bearing in mind that the results, at any rate for a large number of groups, apply equally well to the case of an odd number of groups.

It is necessary to find the values of k and k' for use in equations (1) and (2). This is most easily done by considering the case, $t = 1$, and finding the exact probability under H_0 and H_1 by simple enumeration. We find that

$$k = 2p^{r_1}q^{r_1}, \tag{3}$$

$$k' = p_1^{r_1}q_2^{r_2}\left[\frac{P}{p_2} + \frac{Q}{q_1}\right]. \tag{4}$$

3. To test a hypothesis H_0 against a hypothesis H_1, with a risk α of rejecting H_0 when it is true and β of rejecting H_1 when it is true, the sequential test procedure consists of calculating at each stage of the sampling the likelihood ratio L.

If

$$L > \frac{1-\beta}{\alpha}, \tag{5}$$

we accept H_1, whereas if

$$L < \frac{\beta}{1-\alpha}, \tag{6}$$

we accept H_0, and if L lies in between the values in (5) and (6) we continue to take observations.

The likelihood ratio in this case is

$$L = \frac{\Pr\{t \mid r_1, r_2, H_1\}}{\Pr\{t \mid r_1, r_2, H_0\}}$$

$$= \frac{p_1^{r_1}q_2^{r_2}\left[\dfrac{P}{p_2} + \dfrac{Q}{q_1}\right]}{2p^{r_1}q^{r_2}}\left(\frac{p_2q_1}{q_2p_1}\right)^t. \tag{7}$$

This is a perfectly general expression and valid for all values of the variables. We now propose to consider the particular case where under H_0, $p = q = 0{\cdot}5$. Under H_1 we will take $P = Q = 0{\cdot}5$ and p_1 equal to some particular value other than $0{\cdot}5$. It will be noticed that this simplification has practical application in the consideration of runs above and below the median value in a sequence. It is important that the value used as median should be near to the true value; if it is not, the test instead of picking out a departure from randomness may only be drawing attention to a large difference in the proportion of 1's and 0's. From

§2 we have that $p_2 = q_1$ because $P = Q$ and hence $p_1 = q_2$ since $p_1 + q_1 = 1$. The expression for L may now be re-written under these simplifying assumptions as

$$L = \frac{\tfrac{1}{2}p_1^R\left[\dfrac{1}{q_1}+\dfrac{1}{q_1}\right]}{2(\tfrac{1}{2})^R}\left(\frac{q_1}{p_1}\right)^{2t}$$

$$= 2^{R-1}p_1^{R-2t}q_1^{2t-1}$$

$$= (2p_1)^{R-2t}(2-2p_1)^{2t-1}, \tag{8}$$

where $R = r_1 + r_2$.

Thus L, in this particular case, is seen to be independent of the actual values of r_1 and r_2 and only dependent on the total length of the sequence.

4. Under the sequential system observations are taken for so long as L satisfies the inequality

$$\beta/(1-\alpha) < L < (1-\beta)/\alpha.$$

Expressing this inequality in terms of logarithms and substituting for L the expression obtained in (8), we get that observations are to be taken for so long as

$$\log\frac{\beta}{1-\alpha} < (R-2t)\log 2p_1 + (2t-1)\log(2-2p_1) < \log\frac{1-\beta}{\alpha},$$

which in turn may be written as

$$\log\left\{\frac{\beta}{1-\alpha}(2-2p_1)\right\} < R\log 2p_1 + 2t\log\frac{1-p_1}{p_1} < \log\left\{\frac{1-\beta}{\alpha}(2-2p_1)\right\}. \tag{9}$$

The test procedure may now be carried out graphically. The horizontal axis of the graph corresponds to R and the vertical axis to t. Then if we draw the relationship given in (9), replacing the inequality signs by equality signs in turn, we obtain two parallel straight lines. Then, provided the point (R, t) falls inside these two lines, a further observation is to be taken. If the point falls below the lower line then we accept the alternative hypothesis, H_1, if p_1 is greater than 0·5 or the null hypothesis, H_0, if p_1 is less than 0·5. Similarly if the point falls above the upper line we accept H_0 if p_1 is greater than 0·5 and H_1 if p_1 is less than 0·5.

5. The whole procedure is best illustrated by means of an example. The data are taken from Brunt (1925) and consist of the annual rainfall in inches at London for each of the years 1813–1912. To carry out the test we have described it is necessary to decide how to make the division of the rainfalls into two groups. After 16 years the median rainfall, that is, the average of the eighth and ninth rainfalls when the 16 are ranked in order, is 23·87 in. In Table 1 each year is given the symbol 1 if the rainfall is above this median value and the symbol 0 if it is below the median value. Only the first 45 years are given in Table 1 for economy of space.

In this example we will test the hypothesis of randomness ($p = q = 0{\cdot}5$) against the alternative that $p_1 < 0{\cdot}5$, taking as the particular alternative required to define the procedure, $p_1 = 0{\cdot}3$. Thus we are testing whether wet years tend to be followed by dry years or not. Both α and β will be taken as 0·05, and hence our two equations for the critical lines are

$$R\log(0{\cdot}6) + 2t\log(\tfrac{7}{3}) = \log(\tfrac{1}{19})(1{\cdot}4),$$

$$R\log(0{\cdot}6) + 2t\log(\tfrac{7}{3}) = \log(19)(1{\cdot}4).$$

Evaluating these equations numerically we obtain the two parallel lines

$$R-3\cdot3174t-5\cdot1054=0, \tag{10}$$

$$R-3\cdot3174t+6\cdot4228=0. \tag{11}$$

The lines given by (10) and (11) have been plotted in Fig. 1 and the figure completed by putting in the observed points (R,t) for R greater than 16. Note that the first sixteen observations contain six groups of 0's and six groups of 1's, so that $R=16$, $2t=12$. The chart therefore starts with a point having co-ordinates (16, 6). We see that the point remains inside the lines until R is equal to 36 when $t=13$ and it goes outside the upper line. Thus at this stage we would cease taking observations and accept the hypothesis that wet years tend to be followed by dry years and vice versa.

Table 1. *Rainfall at London in inches*

Year	Rainfall	Year	Rainfall	Year	Rainfall
1813	23·56 0	1828	27·88 1	1843	25·85 1
1814	26·07 1	1829	25·32 1	1844	22·65 0
1815	21·86 0	1830	25·08 1	1845	22·75 0
1816	31·24 1	1831	27·76 1	1846	26·36 1
1817	23·65 0	1832	19·82 0	1847	17·70 0
1818	23·88 1	1833	24·78 1	1848	29·81 1
1819	26·41 1	1834	20·12 0	1849	22·93 0
1820	22·67 0	1835	24·34 1	1850	19·22 0
1821	31·69 1	1836	27·42 1	1851	20·63 0
1822	23·86 0	1837	19·44 0	1852	35·34 1
1823	24·11 1	1838	21·63 0	1853	25·89 1
1824	32·43 1	1839	27·49 1	1854	18·65 0
1825	23·26 0	1840	19·43 0	1855	23·06 0
1826	22·57 0	1841	31·13 1	1856	22·21 0
1827	23·00 0	1842	23·09 0	1857	22·18 0

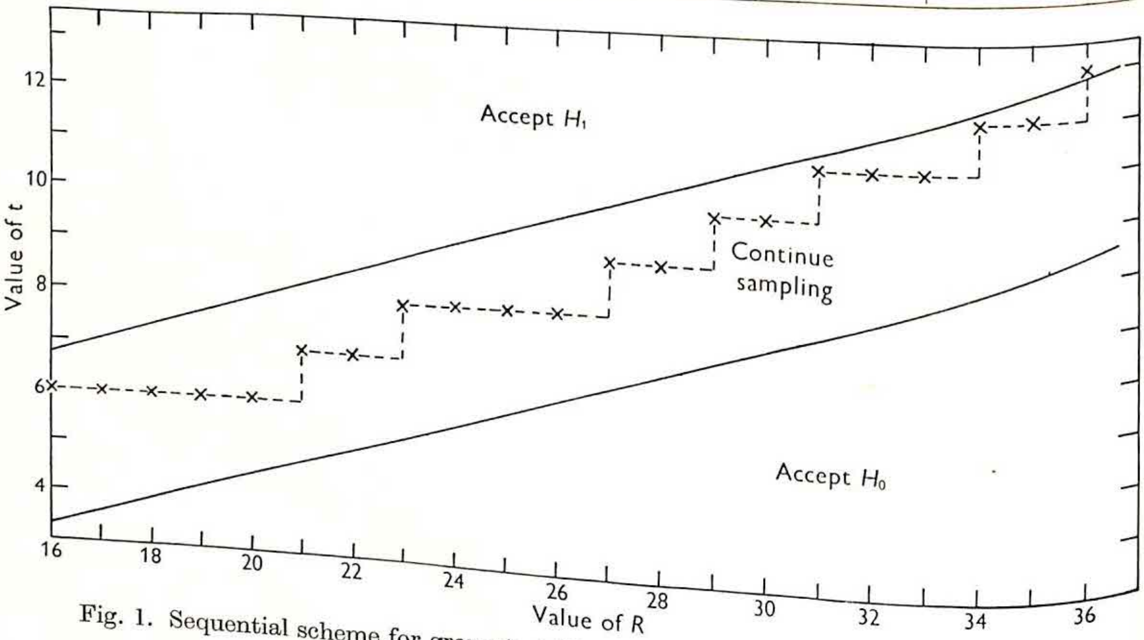


Fig. 1. Sequential scheme for group test H_0: $p=q=0\cdot5$; H_1: $P=Q=0\cdot5$, $p_1=0\cdot3$.

We have already noted that since H_0 states $p = q = 0\cdot 5$, it is essential that as the sequence progresses a check is made that we have about equal numbers of 1's and 0's. This is to prevent the discriminating power of the test being blunted by a wrong choice of the median. In this particular case we find that after thirty-two observations, for instance, there are about equal numbers, in fact seventeen 1's and fifteen 0's, so that there is no reason to question the choice of critical value taken for the median. Had there been a very wide discrepancy it would have been best to have obtained a new median value and recalculated the value of t. The critical lines would remain, of course, the same as before.

In the procedure just carried out we tested the hypothesis of randomness against the hypothesis that $p_1 = 0\cdot 3$, meaning that wet years tend to follow dry years or that a negative correlation exists between successive years. A more general problem would be to test the hypothesis of randomness against a Markoff chain that was either positive or negative in its correlation between successive events. This means that there would be three alternatives, (*a*) positive correlation, (*b*) randomness and (*c*) negative correlation between successive events. In this case the problem could be tackled by setting up two simple sequential schemes. The first would be the same as that already described, and the second scheme would be set up with $p_1 = 0\cdot 7$ (say). If at any stage in the sampling the first scheme told us to accept the hypothesis that $p_1 = 0\cdot 3$ we would do so, and similarly if the second scheme told us to accept $p_1 = 0\cdot 7$ we would do so. However, we would only accept the hypothesis of randomness if both schemes told us to do so. If none of these three results has occurred we continue sampling. Both sets of sequential limits may be drawn on the one diagram, which makes the whole procedure relatively simple to carry out. An example of having two sequential schemes superimposed on the one figure has been given by Armitage (1947). He was concerned with testing whether the mean of a sampled population was equal to some specified value against the alternative that it was either larger or smaller than the specified value.

6. It is useful to have a knowledge of the expected sample size associated with any sequential test before starting the test procedure. Unfortunately, as we are discussing a case where the observations are not independent of one another the usual formulae for estimating the average sample size are not applicable. It is, however, true to say that the sequential procedure will lead to a saving of observations in many cases, since frequently decisions are made on fewer observations than with the fixed sample size of the usual methods of testing.

The test that has been derived and illustrated here has been concerned with the case of $p = q = 0\cdot 5$. It is clear from the derivation in §3 that the procedure could be used for p unequal to q by suitable adjustments. It would also be possible to base the test on a Markoff chain principle under which the observation depends not just on the immediately preceding observation but on the two preceding observations. This case will, it is hoped, be investigated in a later paper.

REFERENCES

ARMITAGE, P. (1947). *J. R. Statist. Soc.* B, **9**, 250.
BRUNT, D. (1925). *Phil. Trans.* A, **225**, 247.
DAVID, F. N. (1947). *Biometrika*, **34**, 335.
FRÉCHET, M. (1938). *Recherches théoriques modernes sur le calcul des probabilités*, Book 2.
MOORE, P. G. (1949). *Biometrika*, **36**, 305.

ON THE MEAN SUCCESSIVE DIFFERENCE AND ITS RATIO TO THE ROOT MEAN SQUARE

By A. R. KAMAT

University College, London

1. Introduction

Let x_i $(i = 1, 2, \ldots, n)$ denote a sequence of n normal variates with means μ_i and a common variance σ^2. If all means μ_i are identical, i.e. if $\mu_i = \mu$, then the best estimator of σ^2 is of course $s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}$. Situations, however, arise in which μ_i may have a 'slow-moving' continuous trend, and for such situations two alternative estimators of σ have been considered. They are: the mean-square successive difference

$$\delta^2 = \sum_{i=1}^{n-1} \frac{(x_i - x_{i+1})^2}{n-1},$$

and the mean successive difference

$$d = \sum_{i=1}^{n-1} \frac{|x_i - x_{i+1}|}{n-1}.$$

Von Neumann, Kent, Bellinson & Hart (1941) have given a brief historical review of these statistics and their use in ballistics and in astronomy, but their results deal only with the statistic δ^2. As to results about d, Arley & Hald (1950) and Guest (1951) have derived the variance of d and have discussed its efficiency.* Quite recently, also, the use of d in quality control chart technique has been discussed by Keen & Page (1953).

Using the results on the absolute moments of the multivariate normal distribution obtained by Nabeya (1951, 1952), and also derived for this purpose by Kamat (1953*a*), the first four moments of d can now be evaluated. In the present paper we propose to discuss the approximate distributions of the mean successive difference, d, and its ratio to the root mean square, viz. d/s, based on their first four moments.

2. The distribution of the mean successive difference

2·1. *Preliminary results*

Let us assume that $\mu_i = \mu$ so that $x_1, x_2, \ldots, x_n$ is a sample from a normal population with constant mean μ and standard deviation σ. We shall use the following notation:

$$\left.\begin{aligned} &\delta_i = x_i - x_{i+1}, \quad d_i = |\delta_i| = |x_i - x_{i+1}| \quad (i = 1, 2, \ldots, n-1), \\ &D = \sum_{i=1}^{n-1} d_i = \sum_{i=1}^{n-1} |x_i - x_{i+1}|. \\ &\text{Then the mean successive difference is} \\ &d = \frac{D}{n-1} = \sum_{i=1}^{n-1} \frac{d_i}{n-1} = \sum_{i=1}^{n-1} \frac{|x_i - x_{i+1}|}{n-1}. \end{aligned}\right\} \quad (1)$$

---

* I understand that similar results (unpublished) were independently obtained by E. C. Fieller & H. O. Hartley.

It is clear that δ_i are normally distributed with zero mean and that

$$\left.\begin{aligned}\sigma(\delta_i) = \sqrt{2}\,\sigma, \quad \rho(\delta_i, \delta_{i+p}) &= -\tfrac{1}{2} \quad (p=1)\\ &= \quad 0 \quad (p>1).\end{aligned}\right\} \tag{2}$$

To find the first four moments of D (and therefore of d), it is necessary to evaluate the following expectations:

$$\left.\begin{array}{llllll}\mathscr{E}(d_1), & & & & & \\ \mathscr{E}(d_1^2), & \mathscr{E}(d_1 d_2), & & & & \\ \mathscr{E}(d_1^3), & \mathscr{E}(d_1^2 d_2), & \mathscr{E}(d_1 d_2 d_3), & & & \\ \mathscr{E}(d_1^4), & \mathscr{E}(d_1^3 d_2), & \mathscr{E}(d_1^2 d_2^2), & \mathscr{E}(d_1^2 d_2 d_3), & \mathscr{E}(d_1 d_2^2 d_3), & \mathscr{E}(d_1 d_2 d_3 d_4).\end{array}\right\} \tag{3}$$

These are either the ordinary moments or the absolute moments of the multivariate normal distribution of $\delta_1, \delta_2, \delta_3, \delta_4$. All of them except the last, viz. $\mathscr{E}(d_1 d_2 d_3 d_4)$, can be evaluated from the results given by Nabeya (1951, 1952) and Kamat (1953*a*) for the bivariate and the trivariate absolute moments, and are as follows:

$$\left.\begin{array}{l}\mathscr{E}(d_1) = \dfrac{2}{\sqrt{\pi}}\sigma, \\ \mathscr{E}(d_1^2) = 2\sigma^2, \quad \mathscr{E}(d_1 d_2) = \left(\dfrac{2\sqrt{3}}{\pi} + \dfrac{1}{3}\right)\sigma^2, \\ \mathscr{E}(d_1^3) = \dfrac{8}{\pi^{\frac{3}{2}}}\sigma^3, \quad \mathscr{E}(d_1^2 d_2) = \dfrac{5}{\sqrt{\pi}}\sigma^3, \quad \mathscr{E}(d_1 d_2 d_3) = \dfrac{4}{\pi^{\frac{3}{2}}}\left(\sqrt{2} + \dfrac{5\pi}{4} - 3\cos^{-1}\dfrac{1}{\sqrt{3}}\right)\sigma^3, \\ \mathscr{E}(d_1^4) = 12\sigma^4, \quad \mathscr{E}(d_1^3 d_2) = \left(2 + \dfrac{9\sqrt{3}}{\pi}\right)\sigma^4, \quad \mathscr{E}(d_1^2 d_2^2) = 6\sigma^4, \\ \mathscr{E}(d_1^2 d_2 d_3) = \left(\dfrac{2}{3} + \dfrac{5\sqrt{3}}{\pi}\right)\sigma^4, \quad \mathscr{E}(d_1 d_2^2 d_3) = \dfrac{12}{\pi}\sigma^4.\end{array}\right\} \tag{4}$$

As regards $\mathscr{E}(d_1 d_2 d_3 d_4)$, two methods were used to find its value. The first is to use the integral form given by Nabeya* (1952). The actual evaluation from his formula of this absolute moment, for the four-variate case with a general correlation pattern, involves considerable difficulty. Even in our particular case, where three of the six correlations are zero and the remaining three are equal, the evaluation is quite elaborate. Using this method,

$$\mathscr{E}(d_1 d_2 d_3 d_4) = \frac{4\sigma^4}{\pi^4} \iiiint_{-\infty}^{\infty} \left\{\frac{\partial^4}{\partial t_1 \partial t_2 \partial t_3 \partial t_4} \left(\exp\left[-\tfrac{1}{2}\Sigma t_i^2 - \Sigma \rho_{ij} t_i t_j\right]\right)\right\} \frac{dt_1\, dt_2\, dt_3\, dt_4}{t_1 t_2 t_3 t_4}, \tag{5}$$

where $i, j = 1, 2, 3, 4$; $i \neq j$ and $\rho_{12} = \rho_{23} = \rho_{34} = -\frac{1}{2}$, $\rho_{13} = \rho_{24} = \rho_{14} = 0$. Carrying out the differentiations in (5), the expression can be reduced to a sum of *nineteen* integrals of the type

$$\iiiint_{-\infty}^{\infty} t_1^l t_2^m t_3^n t_4^p \exp\left[-\tfrac{1}{2}\Sigma t_i^2 - \Sigma \rho_{ij} t_i t_j\right] dt_1\, dt_2\, dt_3\, dt_4, \tag{6}$$

---

* We are indebted to Mr S. Nabeya for sending us the manuscript of his paper (1952) which contains this formula in advance of publication.

where $-1 \leqslant l, m, n, p \leqslant +2$ and $l+m+n+p = 0, -2, -4$. Eighteen of these integrals can be evaluated analytically. For instance,

$$\iiiint\limits_{-\infty}^{\infty} t_1^{-1} t_2^2 t_4^{-1} \exp\left[-\tfrac{1}{2}\Sigma t_i^2 - \Sigma \rho_{ij} t_i t_j\right] dt_1\, dt_2\, dt_3\, dt_4$$

$$= 2 \iint\limits_{0}^{-\frac{1}{2}} d\rho_{12}\, d\rho_{34} \left[\frac{\partial^2}{\partial b_2\, \partial \rho_{23}} \iiiint\limits_{-\infty}^{\infty} \exp\left[-\tfrac{1}{2}\Sigma b_i t_i^2 - \Sigma \rho_{ij} t_i t_j\right] dt_1\, dt_2\, dt_3\, dt_4\right]_{b_2=1,\ \rho_{23}=-\frac{1}{2}},$$

where $$b_1 = b_3 = b_4 = 1, \quad \rho_{13} = \rho_{24} = \rho_{14} = 0.$$

This is further simplified as

$$8\pi^2 \iint\limits_{0}^{-\frac{1}{2}} d\rho_{12}\, d\rho_{34} \left[\frac{\partial^2}{\partial b_2\, \partial \rho_{23}} (A^{-\frac{1}{2}})\right]_{b_2=1,\ \rho_{23}=-\frac{1}{2}}, \quad \text{where} \quad A = (b_2 - \rho_{12}^2)(1-\rho_{34}^2) - \rho_{23}^2$$

$$= 6\pi^2 \iint\limits_{0}^{-\frac{1}{2}} (1-\rho_{34}^2)\left(\tfrac{3}{4} - \rho_{34}^2 - \rho_{12}^2 + \rho_{34}^2 \rho_{12}^2\right)^{-\frac{5}{2}} d\rho_{12}\, d\rho_{34}$$

$$= 4\pi^2 \left(\tfrac{7}{15}\sqrt{5} + \tfrac{8}{9}\sqrt{3} \tan^{-1} \frac{1}{\sqrt{15}}\right). \tag{7}$$

The integral (6) with $l = m = n = p = -1$, however, leads to

$$I = \int_0^{\frac{1}{2}} \sin^{-1}\left(\sqrt{\left(\frac{1-x^2}{3-4x^2}\right)}\right) \frac{dx}{\sqrt{(1-x^2)}},$$

which had to be evaluated by numerical quadrature, and its value was found to be $0{\cdot}328988$. Substituting in (5) the values of all these nineteen integrals, it reduces to

$$\mathscr{E}(d_1 d_2 d_3 d_4) = \frac{16\sigma^4}{\pi^2}\left\{\tfrac{1}{4}\sqrt{5} + \tfrac{1}{2}\sin^{-1}\tfrac{2}{3} + \frac{\sqrt{3}}{2}\sin^{-1}\sqrt{\tfrac{3}{8}} - \tfrac{1}{2}\sin^{-1}\frac{1}{\sqrt{6}} + \tfrac{1}{4}I\right\}$$

$$= 16\sigma^4(1{\cdot}366622)/\pi^2 = (2{\cdot}215484)\,\sigma^4, \tag{8}$$

correct to at least five places of decimals and, possibly, to six places.

This result was checked by finding $\mathscr{E}(d_1 d_2 d_3 d_4)$ using the expansion in series of powers of correlation coefficients given by the present author (1953*a*, pp. 28–30), viz.

$$\mathscr{E}(d_1 d_2 d_3 d_4) = \frac{16\sigma^4}{\pi^2}\{1 + \tfrac{1}{2}\Sigma\rho_{ij}^2 + \Sigma\rho_{ij}\rho_{ik}\rho_{jk} + \tfrac{1}{24}\Sigma\rho_{ij}^4 - \tfrac{1}{4}\Sigma\rho_{ij}^2\rho_{ik}^2$$
$$+ \tfrac{1}{4}\Sigma\rho_{ij}^2\rho_{kl}^2 + \Sigma\rho_{ij}\rho_{kl}\rho_{ik}\rho_{jl} + \ldots\}, \tag{9}$$

where i, j, k, l are all different and assume values 1, 2, 3, 4. For instance, expanding this series up to the tenth power of correlation coefficients and substituting $\rho_{12} = \rho_{23} = \rho_{34} = -\frac{1}{2}$, $\rho_{13} = \rho_{24} = \rho_{14} = 0$, we get

$$\mathscr{E}(d_1 d_2 d_3 d_4) = 16\sigma^4(1{\cdot}366574)/\pi^2 = (2{\cdot}215406)\,\sigma^4, \tag{10}$$

which coincides with the result above to first four significant figures.

2·2. *Moments of d*

Denoting $n-1$ by m,

$$\left.\begin{aligned}
\mathscr{E}(D) &= m\mathscr{E}(d_1),\\
\mathscr{E}(D^2) &= m\mathscr{E}(d_1^2)+2(m-1)\,\mathscr{E}(d_1d_2)+(m-1)\,(m-2)\,\mathscr{E}^2(d_1),\\
\mathscr{E}(D^3) &= m\mathscr{E}(d_1^3)+6(m-1)\,\mathscr{E}(d_1^2d_2)+3(m-1)\,(m-2)\,\mathscr{E}(d_1^2)\,\mathscr{E}(d_1)+6(m-2)\,\mathscr{E}(d_1d_2d_3)\\
&\quad+6(m-2)\,(m-3)\,\mathscr{E}(d_1d_2)\,\mathscr{E}(d_1)+(m-2)\,(m-3)\,(m-4)\,\mathscr{E}^3(d_1),\\
\mathscr{E}(D^4) &= m\mathscr{E}(d_1^4)+8(m-1)\,\mathscr{E}(d_1^3d_2)+4(m-1)\,(m-2)\,\mathscr{E}(d_1^3)\,\mathscr{E}(d_1)+6(m-1)\,\mathscr{E}(d_1^2d_2^2)\\
&\quad+3(m-1)\,(m-2)\,\mathscr{E}^2(d_1^2)+24(m-2)\,\mathscr{E}(d_1^2d_2d_3)+12(m-2)\,\mathscr{E}(d_1d_2^2d_3)\\
&\quad+24(m-2)\,(m-3)\,\mathscr{E}(d_1^2d_2)\,\mathscr{E}(d_1)+12(m-2)\,(m-3)\,\mathscr{E}(d_1d_2)\,\mathscr{E}(d_1^2)\\
&\quad+6(m-2)\,(m-3)\,(m-4)\,\mathscr{E}(d_1^2)\,\mathscr{E}^2(d_1)+24(m-3)\,\mathscr{E}(d_1d_2d_3d_4)\\
&\quad+24(m-3)\,(m-4)\,\mathscr{E}(d_1d_2d_3)\,\mathscr{E}(d_1)+12(m-3)\,(m-4)\,\mathscr{E}^2(d_1d_2)\\
&\quad+12(m-3)\,(m-4)\,(m-5)\,\mathscr{E}(d_1d_2)\,\mathscr{E}^2(d_1)\\
&\quad+(m-3)\,(m-4)\,(m-5)\,(m-6)\,\mathscr{E}^4(d_1).
\end{aligned}\right\}\quad(11)$$

Substituting the values of the various terms, the results (11) lead us to the following central moments of d:

$$\left.\begin{aligned}
\mu_1 &= \frac{2}{\sqrt{\pi}}\sigma = 1{\cdot}128379\sigma,\\
\mu_2 &= \left\{\left(\frac{8}{3}+\frac{4\sqrt{3}-12}{\pi}\right)\frac{1}{m}-\left(\frac{2}{3}+\frac{4\sqrt{3}-8}{\pi}\right)\frac{1}{m^2}\right\}\sigma^2\\
&= \{1{\cdot}052264m^{-1}-0{\cdot}325504m^{-2}\}\,\sigma^2,\\
\mu_3 &= \left\{\left(16\pi+160+24\sqrt{2}-96\sqrt{3}-72\cos^{-1}\frac{1}{\sqrt{3}}\right)\frac{1}{\pi^{\frac{3}{2}}m^2}\right.\\
&\qquad\left.-\left(42\pi+192+48\sqrt{2}-144\sqrt{3}-144\cos^{-1}\frac{1}{\sqrt{3}}\right)\frac{1}{\pi^{\frac{3}{2}}m^3}\right\}\sigma^3\\
&= \{1{\cdot}642670m^{-2}-0{\cdot}870598m^{-3}\}\,\sigma^3,\\
\mu_4 &= \left\{3\left(\frac{8}{3}+\frac{4\sqrt{3}-12}{\pi}\right)^2\frac{1}{m^2}+\left[16\left(180\cos^{-1}\frac{1}{\sqrt{3}}-60\sqrt{2}+210\sqrt{3}-309\right)\right.\right.\\
&\qquad\left.-8\pi(20\sqrt{3}+74)-\frac{16\pi^2}{3}+24\pi^2\alpha\right]\frac{1}{\pi^2m^2}\\
&\qquad+\left[16\left(144\sqrt{2}+468-432\cos^{-1}\frac{1}{\sqrt{3}}-360\sqrt{3}\right)+8\pi(21\sqrt{3}+256)\right.\\
&\qquad\left.\left.+4\pi^2-72\alpha\pi^2\right]\frac{1}{\pi^2m^4}\right\}\sigma^4,\quad\text{where}\quad \alpha=\mathscr{E}(d_1d_2d_3d_4)/\sigma^4=2{\cdot}215484\\
&= \{3(1{\cdot}052264)^2\,m^{-2}+1{\cdot}119816m^{-3}-2{\cdot}040188m^{-4}\}\,\sigma^4.
\end{aligned}\right\}\quad(12)$$

2·3. *Approximate distribution of d*

It is clear from (12) that $\beta_1\to 0$ and $\beta_2\to 3$ as m (and hence n) tends to infinity. Table 1 gives the values of $\sigma(d)$, β_1 and β_2 for $n = 3\,(1)\,10, 15, 20, 25, 30, 40, 50$. The β_1, β_2 points lie in the Pearson Type I region, just above the Type III line. If we may assume that a Pearson Type I curve with the correct first four moments will represent the distribution of d adequately, it is relatively easy to obtain approximations to certain percentage points. The upper and lower 5 and 0·5 % points for the appropriate β_1, β_2 values were found by interpolating in Pearson & Merrington's (1951) tables of standardized percentage points. For $n = 4\,(1)\,8$, the 2·5 and 1 % points were obtained by interpolation in Thompson's (1941*a*) tables of percentage points of the Incomplete Beta-function, first finding the appropriate $p = \frac{1}{2}\nu_2$ and $q = \frac{1}{2}\nu_1$ from the standard relations connecting the Type I parameters with the β_1, β_2 constants of the distribution. Finally, for $n > 8$ it was found that a Type III curve with the correct first three moments was indistinguishable for our purpose from the Type I, and the 2·5 and 1 % points were therefore found with the help of the tables of percentage points of the χ^2-distribution (Thompson, 1941*b*).

Table 1. *Standard deviation,* β_1 *and* β_2 *values for* $d/\sigma = \sum_{i=1}^{n} |x_i - x_{i+1}|/\{(n-1)\,\sigma\}$

n	S.D.	β_1	β_2	n	S.D.	β_1	β_2
3	0·6669	1·0356	4·2962	15	0·2711	0·1638	3·2028
4	·5609	0·7253	3·8940	20	·2334	·1210	3·1499
5	·4927	·5547	3·6857	25	·2080	·0959	3·1188
6	·4443	·4484	3·5548	30	·1895	·0795	3·0984
7	0·4078	0·3760	3·4655	40	0·1636	0·0592	3·0733
8	·3791	·3237	3·4008	50	·1461	·0471	3·0584
9	·3556	·2841	3·3518				
10	·3360	·2531	3·3135				

Note. Mean $d/\sigma = 1{\cdot}128379$. The fourth significant figure in β_1 and the fifth significant figure in β_2 are not reliable.

Table 2. *Percentage points for the approximate distribution of* d/σ

	Lower				Upper			
n	0·5	1	2·5	5	5	2·5	1	0·5
3*	0·09	0·12	0·19	0·27	2·41	2·75	3·16	3·43
4	·17	·22	·29	·37	2·18	2·44	2·76	2·99
5	·23	·28	·36	·44	2·03	2·26	2·53	2·72
6	·28	·33	·41	·49	1·94	2·14	2·37	2·53
7	0·33	0·37	0·45	0·54	1·86	2·04	2·26	2·40
8	·37	·41	·49	·57	1·81	1·97	2·17	2·30
9	·40	·44	·53	·60	1·76	1·91	2·09	2·21
10	·43	·47	·55	63	1·73	1·86	2·03	2·14
15	0·54	0·58	0·65	0·72	1·60	1·71	1·84	1·93
20	·61	·65	·71	·77	1·53	1·62	1·73	1·80
25	·66	·69	·75	·81	1·49	1·57	1·66	1·72
30	·69	·73	·78	·83	1·45	1·52	1·61	1·66
40	0·75	0·78	0·83	0·87	1·41	1·47	1·54	1·59
50	·78	·81	·86	·90	1·38	1·43	1·49	1·53

* For $n = 3$ all percentage points are exact.

For $n = 3$ all percentage points were calculated from the exact distribution given in § 2·4 below. All these results are combined in Table 2. Certain tests of the accuracy of the approximations are given in the following sections.

2·4. *Comparison in the special case $n = 3$*

No direct comparison between exact results and the Pearson Type approximation is possible except in the rather exceptional case $n = 3$. Here, by linear transformation, the probability integral can be expressed in terms of the bivariate normal integral and can be shown to be (see Kamat, 1953*a*, p. 31).

$$P\{d/\sigma \leqslant d_0\} = 1+2\{p_{-\frac{1}{2}}(\sqrt{2}\,d_0, 0)-p_{\frac{1}{2}}(\sqrt{2}\,d_0, 0)+p_{-\sqrt{\frac{3}{4}}}(\sqrt{(\tfrac{2}{3})}\,d_0, 0)-p_{\sqrt{\frac{3}{4}}}(\sqrt{(\tfrac{2}{3})}\,d_0, 0)\}, \quad (13)$$

where $$p_\rho(h, k) = (2\pi)^{-1}(1-\rho^2)^{-\frac{1}{2}}\int_h^\infty\int_k^\infty \exp\left[-\frac{1}{2(1-\rho^2)}(x^2+y^2-2\rho xy)\right]dx\,dy.$$

This function is tabulated in Karl Pearson's (1931) *Tables for Statisticians and Biometricians*, Part II. The following comparison for the upper and lower percentage points for

$$D = (n-1)\,d = 2d$$

shows that the Pearson Type I fit is adequate for practical purposes*.

	Lower			Upper		
	1	2·5	5	5	2·5	1
Integral (13) Pearson fit	0·24 0·26	0·38 0·40	0·53 0·55	4·82 4·81	5·49 5·47	6·31 6·29

2·5. *Sampling experiment*

An extensive sampling experiment was carried out with the help of Hollerith punched-card equipment. 25,000 Hollerith cards, each bearing a random normal deviate from Wold's (1948) tables of random normal deviates, were available. These cards were sorted into 2500 random samples of 10, X_{ji} $(i = 1, 2, \ldots, 10,\ j = 1, 2, \ldots, 2500)$. On a Hollerith Senior Rolling Tabulator the following tabulation was then carried out. From each of the 2500 samples of 10 the ten progressive sums

$$|X_{j1}-X_{j2}|, \quad |X_{j1}-X_{j2}|+|X_{j2}-X_{j3}|, \quad \ldots, \quad \sum_{i=1}^{10}|X_{ji}-X_{ji+1}|$$

were formed and printed (taking $X_{j,11}$ of the first sample as $X_{j+1,1}$ of the next).

For any sample size $n \leqslant 10$, i.e. $n-1 \leqslant 9$, the particular sum $\sum_{i=1}^{n-1}|X_{ji}-X_{ji+1}|$ provides an independent sample value of D, so that for each of the sample sizes $2, 3, \ldots, 10$ (i.e. $n-1 = 1, 2, \ldots, 9$), 2500 independent values of the successive difference sum, D, were available. The same tabulation can also provide 1250 independent sampling values of D for

---

* Calculations made since this table was sent to Press show equally satisfactory agreement at the lower and upper 0·5 % points.

It may be mentioned here that formulae (16)–(19) are exact and that (18) and (19) are obtainable from the first two moments of the non-central χ^2.

Now, for a 'slow-moving' shift in the mean, $\Sigma(\Delta\theta_i)^2/\sigma^2$ will be considerably less than $\Sigma(\theta_i-\bar{\theta})^2/\sigma^2$, and therefore the formulae show that the bias in estimating σ is considerably greater using s^2 than using δ^2 or d. Again, the increase in the variance of the estimate is also larger in the case of s^2. Two examples of how a slow-moving trend affects the bias in the case of the three statistics are given below.

Example (1). $\theta_i = \mu + (0{\cdot}05)\, i\sigma$ $(i = 1, 2, \ldots, n = 15)$.

The percentage bias in the estimate $d = 0{\cdot}06$.
The percentage bias in the estimate $\delta^2 = 0{\cdot}13$.
The percentage bias in the estimate $s^2 = 5{\cdot}00$.

The bias introduced in estimating σ by s^2 is 40 times larger than that introduced by either d or δ^2.* While it is negligible in the latter it may not be so in the former.

Example (2). $\theta_i = \mu + (0{\cdot}04i + 0{\cdot}004i^2)\,\sigma$ $(i = 1, 2, \ldots, n = 10)$.

The percentage bias in the estimate $d = 0{\cdot}24$.
The percentage bias in the estimate $\delta^2 = 0{\cdot}48$.
The percentage bias in the estimate $s^2 = 6{\cdot}54$.

Even in this quadratic trend, the bias is 14 times larger using s^2.

3. THE DISTRIBUTION OF THE RATIO OF THE MEAN SUCCESSIVE DIFFERENCE TO THE ROOT MEAN SQUARE

3·1. *Moments*

Von Neumann (1941) has discussed the exact distribution of the ratio of the mean-square successive difference to the variance, viz. δ^2/s^2, where

$$\delta^2 = \sum_{i=1}^{n-1} \frac{(x_i - x_{i+1})^2}{n-1} \quad \text{and†} \quad s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}.$$

This ratio has been used to test the independence of successive observations and thus to detect serial correlation. It is possible to use for a similar purpose the ratio†

$$W = \frac{d}{s} = \frac{\displaystyle\sum_{i=1}^{n-1} \frac{|x_i - x_{i+1}|}{n-1}}{\sqrt{\left\{\displaystyle\sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{n}\right\}}}. \tag{20}$$

By proving that W and s are independent it can be shown that W has a property of moments which it possesses in common with a number of similar ratios, viz.

$$\mu_r'(W) = \frac{\mu_r'(d)}{\mu_r'(s)}, \tag{21}$$

---

* It is to be noted that d and δ^2 introduce bias of the same magnitude, since while d estimates σ, δ^2 estimates σ^2.

† The divisor used by von Neumann is n, not $n-1$.

where μ_r' is the moment about zero. [See in this connexion Geary (1936) and von Neumann (1941).] Substituting in (21) the moments about zero of d and s, the first four moments of W about zero are obtained as

$$\left.\begin{aligned}
\mu_1' &= \frac{\dfrac{2}{\sqrt{\pi}}\Gamma\left(\dfrac{m}{2}\right)}{\sqrt{\left(\dfrac{2}{m+1}\right)}\Gamma\left(\dfrac{m+1}{2}\right)} = \frac{1\cdot128379\Gamma\left(\dfrac{m}{2}\right)}{\sqrt{\left(\dfrac{2}{m+1}\right)}\Gamma\left(\dfrac{m+1}{2}\right)},\\
\mu_2' &= \frac{m+1}{m}\left\{\frac{4}{\pi}+\left(\frac{8}{3}+\frac{4\sqrt{3}-12}{\pi}\right)\frac{1}{m}-\left(\frac{2}{3}+\frac{4\sqrt{3}-8}{\pi}\right)\frac{1}{m^2}\right\}\\
&= \frac{m+1}{m}\{1\cdot273240+1\cdot052264m^{-1}-0\cdot325504m^{-2}\},\\
\mu_3' &= \frac{\Gamma\left(\dfrac{m}{2}\right)}{\sqrt{\left(\dfrac{2}{m+1}\right)}\Gamma\left(\dfrac{m+1}{2}\right)}\left\{\frac{8}{\pi^{\frac{3}{2}}}+(16\pi+24\sqrt{3}-72)\frac{1}{\pi^{\frac{3}{2}}m}\right.\\
&\qquad +\left(12\pi+208+24\sqrt{2}-120\sqrt{3}-72\cos^{-1}\frac{1}{\sqrt{3}}\right)\frac{1}{\pi^{\frac{3}{2}}m^2}\\
&\qquad \left.+\left(144\sqrt{3}+144\cos^{-1}\frac{1}{\sqrt{3}}-192-48\sqrt{2}-42\pi\right)\frac{1}{\pi^{\frac{3}{2}}m^3}\right\}\\
&= \frac{\Gamma\left(\dfrac{m}{2}\right)}{\sqrt{\left(\dfrac{2}{m+1}\right)}\Gamma\left(\dfrac{m+1}{2}\right)}\{1\cdot43670+3\cdot56206m^{-1}+0\cdot54079m^{-2}-0\cdot87067m^{-3}\},\\
\mu_4' &= \frac{(m+1)^2}{m(m+2)}\left\{\frac{16}{\pi^2}+(96\sqrt{3}-288+64\pi)\frac{1}{\pi^2 m}\right.\\
&\qquad +\left(\tfrac{64}{3}\pi^2+64\sqrt{3}\pi-80\pi+192\sqrt{2}+2048-1152\sqrt{3}-576\cos^{-1}\frac{1}{\sqrt{3}}\right)\frac{1}{\pi^2m^2}\\
&\qquad +\left(24\alpha\pi^2-\tfrac{16}{3}\pi^2-160\sqrt{3}\pi-928\pi+4032\cos^{-1}\frac{1}{\sqrt{3}}\right.\\
&\qquad \left.+4512\sqrt{3}-6480-1344\sqrt{2}\right)\frac{1}{\pi^2m^3}\\
&\qquad +\left(4\pi^2-72\alpha\pi^2+168\sqrt{3}\pi+2048\pi+2304\sqrt{2}\right.\\
&\qquad \left.\left.+7488-5760\sqrt{3}-6912\cos^{-1}\frac{1}{\sqrt{3}}\right)\frac{1}{\pi^2m^4}\right\}\\
&= \frac{(m+1)^2}{m(m+2)}\{1\cdot62114+8\cdot03870m^{-1}+8\cdot24931m^{-2}-2\cdot80941m^{-3}+2\cdot04141m^{-4}\},
\end{aligned}\right\} \tag{22}$$

where
$$m = n-1 \quad\text{and}\quad \alpha = \mathscr{E}(d_1d_2d_3d_4) = 2\cdot215484.$$

3·2. *Approximate distribution of $W = d/s$*

From the moments about zero given above, μ_1', $\sigma(W)$, β_1 and β_2 were calculated for $n = 5, 10, 15, 20, 25, 30, 40, 50$. They are given in Table 4. (It is to be noted that $\mu_3 < 0$, i.e. there is negative skewness.) It is seen from this table that the β_1, β_2 points lie in the

Pearson Type I region, very close to the symmetrical Type II line. For the criterion $C = 1-(n-1)\,\delta^2/(2ns^2)$, which is linearly related with δ^2/s^2, the Pearson Type II curve fitted extremely well, as can be seen from Young (1941) and Hart (1942). The present author has verified that the 5 % points for δ^2/s^2 itself based on the Type II approximation coincide with those given by Hart (1942) based on the exact distribution within two (and sometimes three) places of decimals for $n \geqslant 10$. It seems probable therefore that a Pearson type curve will give a good fit to the distribution of $W = d/s$ also, for $n \geqslant 10$. Table 5 gives the upper and lower 0·5, 1, 2·5 and 5 % points for W based on the Pearson type approximation. The 5 and 0·5 % points were calculated from the Pearson-Merrington (1951) tables. The 2·5 and 1 % points for $n = 10, 15$ and 20 were calculated with the help of Thompson's (1941*a*) tables of the percentage points for the Incomplete Beta-function, while for the remaining sample sizes ($n \geqslant 25$) the normal approximation, which was sufficient for the purpose, was used. It may be remarked that for positive serial correlation, only the lower significance points are necessary. It is hoped to give some further illustrations of the uses of the distribution of d/σ and d/s in a later paper.

Table 4. *The mean, standard deviation, β_1, β_2 values for $W = d/s$**

n	μ_1'	$\sigma(W)$	β_1	β_2
5	1·3421	0·3061	0·009	2·55
10	1·2229	·2116	·013	2·84
15	1·1890	·1709	·011	2·91
20	1·1730	·1470	·009	2·94
25	1·1637	0·1311	0·007	2·95
30	1·1576	·1194	·006	2·96
40	1·1501	·1030	·005	2·96
50	1·1457	·0919	·004	2·98

Note. $\mu_3 < 0$.

Table 5. *Percentage points for $W = d/s$**

n	Lower				Upper			
	0·5	1	2·5	5	5	2·5	1	0·5
10	0·67	0·71	0·79	0·87	1·56	1·63	1·70	1·73
15	·74	·78	·85	·90	1·47	1·52	1·57	1·60
20	·79	·83	·88	·93	1·41	1·46	1·51	1·54
25	0·82	0·86	0·91	0·95	1·38	1·42	1·47	1·49
30	·85	·88	·92	·96	1·35	1·39	1·44	1·46
40	·88	·91	·95	·98	1·32	1·35	1·39	1·41
50	·91	·93	·97	·99	1·30	1·33	1·36	1·38

* It should be noted that following the definition used by von Neumann we have taken $s^2 = \sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{n}$, i.e. taken a divisor of n, not $n-1$.

3·3. *Comparison of the distributions of d/σ and d/s*

It is of interest to compare the distribution of the ratio criterion d/s with that of the numerator d/σ. This comparison does not reveal the familiar properties of a studentized statistic, as the denominator s, which has been computed from the same sample as the numerator, is not independent of d. In fact, d and s are positively correlated since

$$\mathscr{E}(d\,.\,s) = \mathscr{E}\left(\frac{d}{s}s^2\right) = \mathscr{E}\left(\frac{d}{s}\right)\mathscr{E}(s^2) = \frac{\mathscr{E}(d)\,\mathscr{E}(s^2)}{\mathscr{E}(s)},$$

so that

$$\operatorname{cov}(d, s) = \frac{\mathscr{E}(d)}{\mathscr{E}(s)}\{\mathscr{E}(s^2) - \mathscr{E}^2(s)\} = \frac{\mathscr{E}(d)}{\mathscr{E}(s)}\operatorname{var}(s) \geqslant 0.$$

In consequence of this positive correlation the ratio d/s has a smaller variance than d/σ, as can be seen from Tables 1 and 4, and, since $\mathscr{E}(d/s) > \mathscr{E}(d/\sigma)$, the same holds for their 'relative variances', i.e.

$$\operatorname{var}(d/s)/\mathscr{E}^2(d/s) < \operatorname{var}(d/\sigma)/\mathscr{E}^2(d/\sigma).$$

As $n \to \infty$, $\mathscr{E}(d/s) \to \mathscr{E}(d/\sigma) = \frac{2}{\sqrt{\pi}}$, but

$$\frac{\operatorname{var}(d/s)}{\operatorname{var}(d/\sigma)} \to \frac{4\pi + 6\sqrt{3} - 21}{4\pi + 6\sqrt{3} - 18} = 0{\cdot}395,$$

which is less than 1. It may be added that similar properties hold for the statistics δ^2/σ^2 and δ^2/s^2 mentioned in § 3·1.

As will be seen from Tables 2 and 5, a pair of corresponding upper and lower percentage points of d/σ bracket the corresponding percentage points of d/s.

Finally, I wish to express my sincere thanks to Prof. E. S. Pearson and Dr H. O. Hartley for their help and guidance in the course of these investigations.

REFERENCES

ARLEY, N. & HALD, A. (1950). *Math. Tidsskr.* B, p. 86.
GEARY, R. C. (1936). *Biometrika*, **28**, 295.
GUEST, P. G. (1951). *Suppl. J. Roy. Statist. Soc.* **13**, 233.
HART, B. I. (1942). *Ann. Math. Statist.* **12**, 445.
KAMAT, A. R. (1953*a*). *Biometrika*, **40**, .
KAMAT, A. R. (1953*b*). (Unpublished results.)
KEEN, J. & PAGE, D. J. (1953). *Applied Statistics*, **2**, 13.
NABEYA, S. (1951). *Ann. Inst. Statist. Math.* **3**, 2.
NABEYA, S. (1952). *Ann. Inst. Statist. Math.* **4**, 15.
PEARSON, E. S. & MERRINGTON, M. (1951). *Biometrika*, **38**, 4.
PEARSON, K. (1931). *Tables for Statisticians and Biometricians*, **2**. Biometrika Trust, London.
THOMPSON, C. M. (1941*a*). *Biometrika*, **32**, 151.
THOMPSON, C. M. (1941*b*). *Biometrika*, **32**, 187.
VON NEUMANN, J. (1941). *Ann. Math. Statist.* **12**, 367.
VON NEUMANN, J., KENT, R. H., BELLINSON, H. R. & HART, B. I. (1941). *Ann. Math. Statist.* **12**, 153.
YOUNG, L. C. (1941). *Ann. Math. Statist.* **12**, 293.
WOLD, H. (1948). *Tracts for Computers*, No. 25. Cambridge University Press.

THE EFFECT OF UNEQUAL GROUP VARIANCES ON THE F-TEST FOR THE HOMOGENEITY OF GROUP MEANS

BY G. HORSNELL

1. INTRODUCTION

David & Johnson (1951) have studied briefly, *inter alia*, the effect of unequal group variances on the nominal significance levels of the F-test (variance ratio) when used to investigate differences among a number of group means. The present note takes those calculations a step further and also considers the effect on the power of the test in a special case.

If a total of N observations of a random variable are divided into l groups, with frequencies $n_t\,(t = 1, 2, \ldots, l)$, where $N = \sum_t n_t$, if S_1 and S_2 are the usual *between group* and *within group* sums of squares and if F_α is the 100α % significance level of the F-distribution with $\nu_1 = l-1$ and $\nu_2 = N-l$ degrees of freedom, then the David-Johnson method consists in approximating to the distribution of

$$x = S_1 - \left\{\frac{l-1}{N-l} F_\alpha\right\} S_2. \tag{1}$$

Three types of curve have been used in deriving the approximations:

(i) the Edgeworth series in the form

$$p(X) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}X^2}\left\{1 + \frac{\kappa_3(X)}{3!} H_3(X) + \frac{\kappa_4(X)}{4!} H_4(X) + \frac{10\kappa_3^2(X)}{6!} H_6(X)\right\}, \tag{2}$$

where $X = (x - \kappa_1(x))/\kappa_2^{\frac{1}{2}}(x)$, $\kappa_r(x)$ is the rth cumulant of x and the H-functions are Hermite polynomials;

(ii) Johnson's (1949) unbounded curve S_U, and the log-normal curve;

(iii) a Pearson curve, usually of type IV.

The curves were fitted using the first four moments of x derived from the moments of S_1 and S_2 given by David & Johnson. Apart from unequal group variances, κ_{2t}, the usual assumptions made in the analysis of variance are supposed satisfied.

The present investigation can only be regarded as exploratory in the sense that it uses the general formulae to illustrate certain special cases. The number of groups has been confined to $l = 4$, and the six different sets of values for the group variances $\kappa_{2t}\,(t = 1, 2, 3, 4)$ considered by David & Johnson have been used. These values are as follows:

Set	κ_{21}	κ_{22}	κ_{23}	κ_{24}
A	1	1	1	2
B	1	1	2	2
C	1	2	2	2
D	1	1	1	3
E	1	1	2	3
F	1	1	3	3

2. Effect on significance level when the group frequencies are equal

Four cases with an equal number of observations, n_t, in each group, namely, 5, 10, 15 and 20 respectively, will be considered first. The values obtained for the actual tail areas cut off when using the nominal 5 % and 1 % significance levels of F are given in Table 1. A column for the case of equal variances is also included to give an indication of the accuracy of the method. The values obtained from approximating to the distribution of x by the three types of curve are in fairly close agreement, the Pearson curve, where used, nearly always giving a result closer to that of the S_U than to that of the Edgeworth series. There does not appear to be any change in the error of the significance level (which in every case is small) as the sample size increases.

Table 1. *Equal groups*

(*a*) 5 % *nominal significance level*

n_t	Curve fitted	Equal variances	Unequal variances					
			Set A	Set B	Set C	Set D	Set E	Set F
5	Edgeworth	0·051	0·056	0·056	0·054	0·063	0·060	0·062
	S_U	·052	·058	·058	·055	·069	·064	·066
	Type IV	—	—	—	—	·068	—	—
10	Edgeworth	0·053	0·059	0·058	0·056	0·064	0·062	0·063
	S_U	·050	·055	·055	·052	·063	·059	·061
	Type IV	·050	—	—	—	·063	—	—
15	Edgeworth	0·053	0·058	0·057	0·056	0·062	0·061	0·061
	S_U	·049	·054	·054	·052	·062	·058	·059
	Type IV	—	—	—	—	·062	—	—
20	Edgeworth	0·053	0·057	0·057	0·055	0·060	0·059	0·060
	S_U	·049	·054	·054	·052	·061	·058	·059
	Type IV	·049*	—	—	—	·059	—	—

* Value obtained from Pearson type V curve as the β_1, β_2 point for x lay close to the type V line.

(*b*) 1 % *nominal significance level*

n_t	Curve fitted	Equal variances	Unequal variances					
			Set A	Set B	Set C	Set D	Set E	Set F
5	S_U	0·010	0·013	0·012	0·011	0·018	0·015	0·016
10	S_U	·010	·013	·013	·011	·018	·015	·016
15	S_U	·010	·012	·012	·011	·017	·014	·015
20	S_U	·009	·012	·012	·011	·016	·014	·015

3. EFFECT ON SIGNIFICANCE LEVEL WHEN THE GROUP FREQUENCIES ARE UNEQUAL

Welch (1937) considered the effect of unequal variances on the t-test for the difference of two population means. He found that for $n_1 = n_2 = 10$ and a nominal significance level of 5 %, the true level lay approximately between 5 and 6·5 % for all values of κ_{21}/κ_{22}. For $n_1 = 5$ and $n_2 = 15$, however, he found that the test could be seriously misleading, and showed that the significance of the difference between the two means would tend to be underestimated when $\kappa_{21} < \kappa_{22}$ and overestimated when $\kappa_{21} > \kappa_{22}$ for a given nominal level of significance. This problem has also been considered by Gronow (1951, 1953).

Table 2 shows that these conclusions can be generalized to the case of more than two groups. The sets of variances considered are set A and set D. The error in the significance level appears to depend only on the ratios of the numbers in the groups, n_t/N, and not on their absolute magnitudes. For the error in the significance level to be a minimum, it is necessary to take slightly more observations in the group with the high variance. An empirical rule seems to be that the ratios should lie about half-way between equality and the ratios of the standard deviations in the groups, for the cases considered here. A similarity may be noted between this condition and that in stratified sampling, where, to obtain a linear estimator of a population parameter with minimum variance, it is necessary to take a sample from each stratum proportional in size to the standard error and the size of the stratum.

4. EFFECT OF UNEQUAL VARIANCES ON THE POWER OF THE TEST

The David-Johnson method will now be used to examine how inequality of variance affects the power of the test, that is, the probability that it will detect real differences between the population group means. Again, it is only possible to examine particular cases. For the standard case of 'normal theory', where $\kappa_{2t} = \sigma^2$ for all t, the power of the test depends on (*a*) $\nu_1 = l-1$ and $\nu_2 = N-l$, the between- and within-group degrees of freedom, and (*b*) on a single parameter which has been termed the non-central parameter, measuring the degree of heterogeneity among the group means. If the population group means are κ_{1t} $(t = 1, 2, \ldots, l)$, if $\kappa_{1.} = \sum_t (n_t\kappa_{1t})/\sum_t n_t$ and if we write

$$C_t = \kappa_{1t} - \kappa_{1.},$$

then Tang (1938), in preparing his tables, used as the non-central parameter,

$$\phi = \{\sum_t n_t C_t^2/(\sigma^2(\nu_1+1))\}^{\frac{1}{2}}. \tag{3}$$

This notation was repeated in their power charts by Pearson & Hartley (1951). We shall write $\beta(\phi)$ for the power associated with the standard conditions. When the within-group variances κ_{2t} differ, several forms of comparison may be of interest, and these are discussed below.

We shall consider only the set of variances D, i.e. with κ_{2t} as (1, 1, 1, 3), and with these we take the four combinations of the group frequencies n_t shown in Table 4 (with $\Sigma n_t = 40$) and two different cases of divergent means, namely,

Case (*i*): $C_2 = C_3 = C_4$ but C_1 different, i.e. there is a divergent mean in a group with low variance.

Case (*ii*): $C_1 = C_2 = C_3$ but C_4 different, i.e. the divergent mean is in the group with high variance.

The first step in the investigation is to confirm that the David-Johnson method provides close approximations to the power when the κ_{2t} are equal. A comparison is made in Table 3

Table 2. *Unequal groups*

(1) *Set A* (κ_{2t}'s: 1, 1, 1, 2). (*Tail areas estimated from* S_U *curves*)

n_1	n_2	n_3	n_4	$\sum_t n_t$	Nominal significance level	
					5 %	1 %
8	8	8	16	40	0·032	0·006
10	10	10	10	40	·055	·013
11	11	12	6	40	·077	·020
16	16	16	32	80	0·030	0·005
20	20	20	20	80	·054	·012
22	23	23	12	80	·076	·020

(2) *Set D* (κ_{2t}'*s*: 1, 1, 1, 3). $\sum_t n_t = 40$

n_1	n_2	n_3	n_4	Nominal significance level			
				5 %			1 %
				Edgeworth	S_U	Type IV	S_U
7	7	7	19	0·021	0·017	0·016	0·002
9	9	9	13	·049	·043	—	·004
9	9	10	12	·054	·049	—	·012
9	10	10	11	0·059	0·056	—	0·015
10	10	10	10	·064	·063	·063	·018
12	12	12	4	·103*	·132	·134	·052

* The tail ordinates of the Edgeworth curve had a 'hump' in this case and would have been negative for part of the range but for the term in $\kappa_3^2(X)$.

(3) *Set D* (κ_{2t}'*s*: 1, 1, 1, 3). $\sum n_t = 80$. (*Tail areas estimated from* S_U *curves*)

n_1	n_2	n_3	n_4	Nominal significance level	
				5 %	1 %
13	13	14	40	0·013	0·002
17	17	17	29	·026	·005
18	18	18	26	·042	·009
19	19	20	22	·053	·014
19	20	20	21	0·057	0·015
20	20	20	20	·061	·016
24	24	24	8	·131	·048

which is clearly satisfactory, and gives some confidence in using the method when the variances are unequal. In the further work, the Edgeworth series has been used, except where indicated by notes below Table 4.

Table 3. *Equal variances.* $\sum_t n_t = 40$. *Comparison of approximations to true significance level and power*

Significance level		Power			
			$\phi = 1$	$\phi = 2$	$\phi = 2{\cdot}5$
True level	0·050	Tang	0·325	0·904	0·987
Edgeworth series	·053	Edgeworth series	·332	·901	·986
S_U	·050	Type III curve	·329	·903	·986
Type IV	·050	Log-normal curve	·328	—	—

We have now to consider what comparison may be of interest in illustrating the effect of unequal variances. In the first place, we may compare the power computed as described for cases (i) and (ii) with that obtained if Tang's tables or the Pearson-Hartley charts are entered with

$$\phi = \phi_1 = \{\sum_t n_t C_t^2/(\sigma_L^2(\nu_1+1))\}^{\frac{1}{2}}, \tag{4}$$

where σ_L^2 is the common low variance of the first three groups (taken as unity in the calculations). In this way we can judge how, in the cases illustrated, the unrecognized presence of a single high group-variance will reduce the chance of detecting a divergent mean value, using the ordinary method of analysis.

An alternative approach is to replace σ^2 in equation (3) by $\kappa_{2.} = \sum_t n_t \kappa_{2t}/N$, i.e. to enter the power function table with

$$\phi = \phi_2 = \{\sum_t n_t C_t^2/(\kappa_{2.}(\nu_1+1))\}^{\frac{1}{2}}. \tag{5}$$

For small differences among the κ_{2t} it would be expected that the use of $\kappa_{2.}$ in (5) would provide a good approximation to the true power. If the power function is to be used, in conjunction with a rough estimate of within-group variability derived from past experience, to explore in advance the size of samples required to make a critical experiment worth while, then this rough estimate may well approximate to the average variance $\kappa_{2.}$. The example we have taken with the κ_{2t} having values $(1, 1, 1, 3)$ is perhaps more extreme than would be expected under these conditions.

Table 4 gives the appropriate comparisons. For both cases (i) and (ii) the C_t were chosen for convenience to make the ϕ_1 of equation (4) assume exact values of 1, 2 and 2·5. For given $\Sigma n_t C_t^2$ and hence fixed ϕ_1 (since $\sigma_L^2 = 1$), ϕ_2 will be constant for a comparison between cases (i) and (ii) within a cell of the table, but will alter with the group frequencies in passing down a column, since $\kappa_{2.}$ is a weighted mean of the κ_{2t}.

Taking first the true power (as calculated from the approximations), we see that within a given cell of the table the true power is always less for case (ii) than for case (i), apart from one exception where the power is small. This means that for a worth-while power, we are less likely to detect a divergent mean when this occurs in a group with large variance. As has been seen already in connexion with Table 2 (2), when n_4 is well above or below the average group frequency of 10, the significance level is seriously affected for this case D.

Consider now comparisons with the power $\beta(\phi)$ obtained from Tang's tables, based on taking $\phi = \phi_1$. We find again, as we should expect, that for a given set of C_t an increase in the variances of one group lowers the power of the test. Except in the case where $\phi_1 = 1$ and the n_t are (12, 12, 12, 4), where the actual significance level is much higher than the nominal 5 %, the power for cases (i) and (ii) is always lower, and generally much lower than $\beta(\phi = \phi_1)$.

Table 4. κ_{2t}'s: 1, 1, 1, 3. $\sum_t n_t = 40$. *Nominal significance level* 5 %

t	n_t	Real significance level using*	Approximate power based on Edgeworth series		
			$\phi = \phi_1 = 1$	$\phi = \phi_1 = 2$	$\phi = \phi_1 = 2{\cdot}5$
1 2 3 4	7 7 7 19	E 0·021 S_U 0·017 Type IV 0·016	Approx. power (i) 0·128 (ii) 0·115 $\beta(\phi = \phi_1 = 1)$ 0·325 $\beta(\phi = \phi_2 = 0{\cdot}716)$ 0·18	Approx. power (i) 0·576 (ii) 0·540 $\beta(\phi = \phi_1 = 2)$ 0·904 $\beta(\phi = \phi_2 = 1{\cdot}432)$ 0·61	Approx. power (i) 0·835 (ii) 0·768 $\beta(\phi = \phi_1 = 2{\cdot}5)$ 0·987 $\beta(\phi = \phi_2 = 1{\cdot}790)$ 0·82
1 2 3 4	9 9 10 12	E 0·054 S_U 0·049 Type IV —	Approx. power (i) 0·211 (ii) 0·201 $\beta(\phi = \phi_1 = 1)$ 0·325 $\beta(\phi = \phi_2 = 0{\cdot}791)$ 0·21	Approx. power (i) 0·724 (ii) 0·644 $\beta(\phi = \phi_1 = 2)$ 0·904 $\beta(\phi = \phi_2 = 1{\cdot}581)$ 0·71	Approx. power (i) 0·918 (ii) 0·830 $\beta(\phi = \phi_1 = 2{\cdot}5)$ 0·987 $\beta(\phi = \phi_2 = 1{\cdot}976)$ 0·89
1 2 3 4	10 10 10 10	E 0·064 S_U 0·063 Type IV 0·063	Approx. power $(\text{i})_1$† 0·216 $(\text{ii})_2$† 0·233 $\beta(\phi = \phi_1 = 1)$ 0·325 $\beta(\phi = \phi_2 = 0{\cdot}816)$ 0·23	Approx. power (i) 0·764 (ii) 0·672 $\beta(\phi = \phi_1 = 2)$ 0·904 $\beta(\phi = \phi_2 = 1{\cdot}633)$ 0·74	Approx. power (i) 0·936 (ii) 0·846 $\beta(\phi = \phi_1 = 2{\cdot}5)$ 0·987 $\beta(\phi = \phi_2 = 2{\cdot}041)$ 0·91
1 2 3 4	12 12 12 4	E 0·103 S_U 0·132 Type IV 0·134	Approx. power (i) 0·365 $(\text{ii})_3$† 0·339 $\beta(\phi = \phi_1 = 1)$ 0·325 $\beta(\phi = \phi_2 = 0{\cdot}913)$ 0·28	Approx. power (i) 0·880 (ii) 0·759 $\beta(\phi = \phi_1 = 2)$ 0·904 $\beta(\phi = \phi_2 = 1{\cdot}826)$ 0·84	Approx. power (i) 0·979 (ii) 0·892 $\beta(\phi = \phi_1 = 2{\cdot}5)$ 0·987 $\beta(\phi = \phi_2 = 2{\cdot}282)$ 0·96

* Three methods of approximating to the real significance level have been used: E, with the Edgeworth series; S_U, with Johnson's (1949) curve; type IV, with the Pearson curve.

† Only the Edgeworth series was used in calculating the power, except in the cases marked $(\text{i})_1$, $(\text{ii})_2$, $(\text{ii})_3$; S_U curves and the log-normal curve were used in addition to the Edgeworth series to calculate the power here, the values obtained being: $(\text{i})_1$ S_U 0·228, $(\text{ii})_2$ S_U 0·245, $(\text{ii})_3$ log-normal 0·354.

However, more instructive comparisons are made taking $\phi = \phi_2$. Here in a number of cases $\beta(\phi = \phi_2)$ lies between the power for case (i) and case (ii). For low values of the power, say less than 0·50, its values may be considerably influenced by a wrong start to the curve at $\phi = 0$, i.e. by the difference between the actual and nominal significance levels. But it will be seen that as the power becomes large so that there is a worth-while chance of establishing the significance of differences in means, the tabular power based on the weighted mean variance gives a very reasonable approximation to the true power, particularly when the divergent mean is in a group with lower variance.

Comparison is made easier by plotting the power curves as in Figs. 1 and 2 using $\phi = \phi_2$ as abscissa. The curves for the frequency combination (9, 9, 10, 12) have not been drawn in as they fall very close to those for equal groups, (10, 10, 10, 10). The standard curve, with ordinate $\beta(\phi)$, is shown in both figures by a broken line.

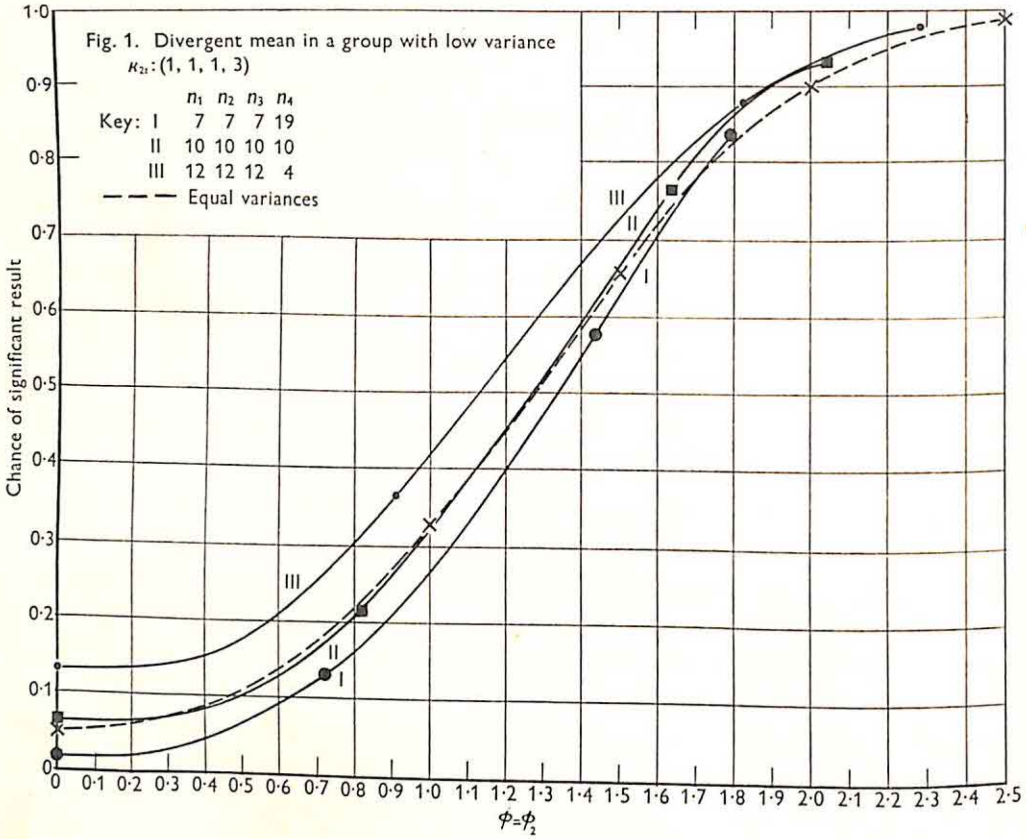


Fig. 1. Divergent mean in a group with low variance
κ_{2i}: (1, 1, 1, 3)
n_1 n_2 n_3 n_4
Key: I 7 7 7 19
II 10 10 10 10
III 12 12 12 4
Equal variances
Chance of significant result
0
0·1
0·2
0·3
0·4
0·5
0·6
0·7
0·8
0·9
1·0
0 0·1 0·2 0·3 0·4 0·5 0·6 0·7 0·8 0·9 1·0 1·1 1·2 1·3 1·4 1·5 1·6 1·7 1·8 1·9 2·0 2·1 2·2 2·3 2·4 2·5
$\phi=\phi_2$
I
II
III

Fig. 1.

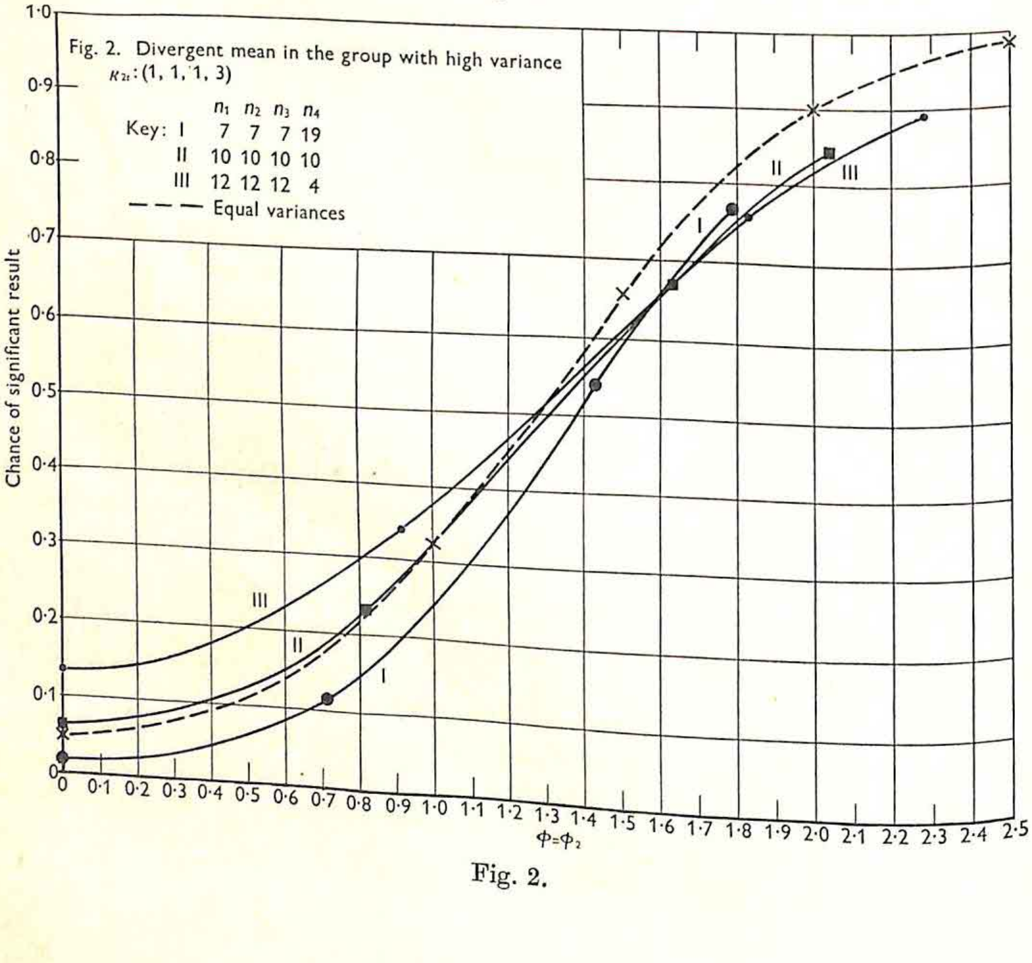


Fig. 2. Divergent mean in the group with high variance
κ_{2i}: (1, 1, 1, 3)
n_1 n_2 n_3 n_4
Key: I 7 7 7 19
II 10 10 10 10
III 12 12 12 4
Equal variances
Chance of significant result
0
0·1
0·2
0·3
0·4
0·5
0·6
0·7
0·8
0·9
1·0
0 0·1 0·2 0·3 0·4 0·5 0·6 0·7 0·8 0·9 1·0 1·1 1·2 1·3 1·4 1·5 1·6 1·7 1·8 1·9 2·0 2·1 2·2 2·3 2·4 2·5
$\phi=\phi_2$
I
II
III

Fig. 2.

For *case* (i) (Fig. 1), the curve for frequencies (10, 10, 10, 10) is very close to the standard. The curves for (7, 7, 7, 19) and (12, 12, 12, 4) lie, respectively, below and above the other two curves for low ϕ, owing to the different values of the significance levels. Allowing for this, however, we see in the different slopes of the curves the advantage of having more observations in the group with high variance.

For *case* (ii) (Fig. 2), the curve for equal frequencies falls away from and below the standard curve as ϕ increases. The curve for frequencies (7, 7, 7, 19) starts at $\phi = 0$ with an ordinate of 0·02 instead of 0·05, but has passed above the equal frequency curve when $\phi = 2$, though it has not caught up with the standard. On the other hand, the curve for frequencies (12, 12, 12, 4), in spite of its high start at 0·13 when $\phi = 0$, shows the lowest power when $\phi = 2$.

It is difficult to summarize these results in any simple form, but perhaps the following is a fair statement of the position. In general, where there is no very clear information as to how any heterogeneity in group variances is apportioned, it will be best to work with equal group frequencies. If, however, there are definite grounds for believing that the observations in one or more groups have a variance above the average, we must avoid taking less than the average number of observations from these groups. This is because we do not want to run the risk of claiming a significant effect (when there are no real differences in means) considerably more often than has been allowed for on the basis of the significance level chosen. It will, indeed, be worth while taking a few more observations from the groups whose variance we believe to be above average, and there is no great danger of overdoing this. For there can be no objection to the risk of wrongly claiming significance being less than we have allowed for (e.g. for the power curve to be below the standard near $\phi = 0$), if the true curve makes a good recovery as $\beta(\phi = \phi_2)$ increases above 0·50.

5. ALTERNATIVE TEST PROCEDURES

Where the group variances are known, the statistic

$$L = \sum_t \omega_t (k_{1t} - \bar{k}_1)^2 \tag{6}$$

may be used to test for the homogeneity of $t = 1, 2, \ldots, l$ group means, where $\omega_t = n_t/\kappa_{2t}$, k_{1t} = the tth group sample mean and $\bar{k}_1 = \sum_t \omega_t k_{1t} / \sum_t \omega_t$.

It may easily be shown that L is distributed as χ^2 with $l-1$ degrees of freedom when the group means are all equal. In another connexion, the L-test has been found useful in comparing estimates of the constants of a number of probit lines (see, for example, Miller, Bliss & Braun, 1939).

When the estimates of the group variances are based on sufficiently large numbers of observations, ω_t may be replaced by $w_t = n_t/k_{2t}$, where k_{2t} $(t = 1, 2, \ldots, l)$ are the estimated variances.

$$L' = \sum_t w_t (k_{1t} - \bar{k}'_1)^2, \tag{7}$$

where $\bar{k}'_1 = \sum_t w_t k_{1t} / \sum_t w_t$, will then be distributed approximately as χ^2 with $l-1$ degrees of freedom, and the approximate power of the test may be obtained from the non-central χ^2 distribution. For small sample sizes the above test will not be very trustworthy. James (1951) has devised corrections to the significance levels of χ^2 depending on the estimates of group variances, while Welch (1951) uses an F-distribution to approximate to the distribution of L'. Both procedures should provide a better test.

The fitting of S_U curves in this investigation was facilitated by formulae due to J. Draper (1952).

The author's thanks are due to Prof. E. S. Pearson, Dr N. L. Johnson and Dr F. N. David for suggesting the topic of this paper and for assistance in its preparation. Acknowledgement is also made to the Chief Scientist, Ministry of Supply, for permission to publish this note.

REFERENCES

David, F. N. & Johnson, N. L. (1951). *Biometrika*, **38**, 43.
Draper, J. (1952). *Biometrika*, **39**, 290.
Gronow, D. G. C. (1951). *Biometrika*, **38**, 252.
Gronow, D. G. C. (1953). *Biometrika*, **40**, 222.
James, G. S. (1951). *Biometrika*, **38**, 324.
Johnson, N. L. (1949). *Biometrika*, **36**, 149.
Miller, L. C., Bliss, C. I. & Braun, H. A. (1939). *J. Amer. Pharm. Ass.* **28**, 644.
Pearson, E. S. & Hartley, H. O. (1951). *Biometrika*, **38**, 112.
Tang, P. C. (1938). *Statist. Res. Mem.* **2**, 126.
Welch, B. L. (1937). *Biometrika*, **29**, 350.
Welch, B. L. (1951). *Biometrika*, **38**, 330.

THE ESTIMATION OF POPULATION PARAMETERS FROM DATA OBTAINED BY MEANS OF THE CAPTURE-RECAPTURE METHOD

III. AN EXAMPLE OF THE PRACTICAL APPLICATIONS OF THE METHOD

By P. H. LESLIE, DENNIS CHITTY and HELEN CHITTY

Bureau of Animal Population, Department of Zoological Field Studies, Oxford

CONTENTS

	PAGE
1. Introduction	137
2. Description of the area and field methods	138
3. Statistical analysis	141
A. The *Microtus* population	142
(i) Preliminary analysis	143
(ii) Data for the marked individuals	146
(iii) Methods of calculating the death-rate and of constructing life-tables	151
B. The *Clethrionomys* population	156
(i) Preliminary analysis	156
(ii) Data for the marked individuals	159
(iii) Estimation of total numbers	160
4. Discussion of results	163
5. Summary	165
Appendix	166
References	169

1. Introduction

The theory of estimating various population parameters from data obtained by the capture-recapture method has been discussed in two earlier papers (Leslie & Chitty, 1951; Leslie, 1952), with special reference to the types of problem which may arise in applying this method to populations of small mammals. It was also shown, by means of an actual sampling experiment carried out on an artificial population of counters, that we should obtain satisfactory estimates of these parameters, together with their variances, from equations which are based mathematically on a deterministic model of the population, provided that the sampling of the individual members is wholly at random and that the size of the samples is not too small. We now propose to examine in some detail the results of an experiment in the field, in which two populations of small rodents, namely, the field vole *Microtus agrestis* and the bank vole *Clethrionomys glareolus*, living on the same area, were sampled by means of a live-trapping technique over a period of very nearly two years, between May 1948 and April 1950.

The most important question which immediately arises in the application of this form of analysis to any set of field data is whether the basic assumptions which are made in the theoretical development were in any way realized. Thus, all the equations which have been given for estimating the different parameters are based on the assumptions that the sampling of the individuals was entirely at random, and that all classes of marked and unmarked animals were being caught with equal facility. It is necessary, therefore, in the first place to examine a set of data with this point particularly in mind, for if there is a complete breakdown in these assumptions, then no valid estimates can be made of the parameters we wish to determine. The results presented here will show how dangerous it may be to apply theoretical methods of estimation to unsuitable data, without some such form of preliminary analysis.

The main purpose of the present paper is, then, to see whether any of the methods of estimation which have been developed on a purely theoretical basis can be applied to a set of field-data obtained by trapping a population of living animals. We shall not be concerned either with the purposes of the original inquiry, or with the biological or ecological interpretations which may be placed on any of the results. These, together with the reasons why certain technical procedures were adopted in the field, we propose to discuss in greater detail elsewhere.

Finally, a word as to nomenclature. It is evident that in trapping a population of wild animals we are not obtaining a sample covering all the possible age-groups alive at a given time. Thus, by this method of sampling, we can have no knowledge of the number of young in the nest, and it is not until these young grow up and become active that they gradually become part of the trappable population. It would be tedious and cumbersome to refer each time to the 'population at risk of capture', and it must be clearly understood that when, for the sake of simplicity, we use instead the term 'total population', or 'the total number of individuals alive at time t', we mean only those members of the population which are capable of entering the traps. In the case of a species with an intermittent breeding season, however, these members may constitute the entire population on the area at certain times of the year. Similarly, for the sake of brevity, we shall use 'dilution-rate' and 'death-rate' as comprehensive terms which would also include any rates of immigration and emigration to and from the area.

Since we shall have frequently to cross-refer to various sections of the two previous papers (Leslie & Chitty, 1951; Leslie, 1952), these references will in future be given in the shortened form I, §x and II, §y respectively.

2. DESCRIPTION OF THE AREA AND FIELD METHODS

The trapping area (Fig. 1) of about 4 acres lay on the north-east side of Lake Vyrnwy, Montgomeryshire, and had been planted in 1947 with Norway spruce, which was about 1 ft. high at the time of our studies. The boundaries of the area were formed by the lake, a stream, and mature plantations or scrub, and a road ran through it from south-east to north-west. The *Microtus* population was therefore almost entirely self-contained, the only open channel for any immigration or emigration being along the road at either end. The mature-plantation and scrub habitats, however, did not prevent other species from moving to and from the trapping area.

The field methods were similar to those used in a previous study (Chitty, 1952), the chief differences being that sampling points were allotted at random and that traps were pre-baited for 48 hr. before being set. We hoped by these procedures to avoid the error, met with before, of drawing samples which were biased by an excess of marked animals (Chitty & Kempson, 1949).

The area was divided into five sections (0, 1, 2, 3, 4), and iron railings along the road formed a convenient baseline to which points within the sections could be referred. Prebaiting points were determined by means of a table of random numbers, each section being sampled in proportion to its area. From May to July 1948 positions were chosen at random within sections, all of which were trapped simultaneously, though lightly, between 9 a.m. and dusk. On each of the following two or three days the traps were lifted and redistributed at sites randomly chosen and already prebaited. In September and October 1948 we still trapped on the same system except that the traps were set overnight. In November 1948, at the suggestion of Mr M. J. R. Healy, a system of stratified random sampling was adopted, the most usual method being to allot either one or two positions at random within each 10 yd. square. If a point chanced to fall beyond the margin of an incomplete square no trap was prebaited, and at points on the area where no prebait was taken, no trap was set; otherwise either one or two traps were used at each point. The area was now trapped in sections on different days, the order being determined at random. From November 1948 to May 1949 traps were set in the afternoon and visited next morning, but from June 1949 they were set during the morning and visited twice within the next 24 hr.

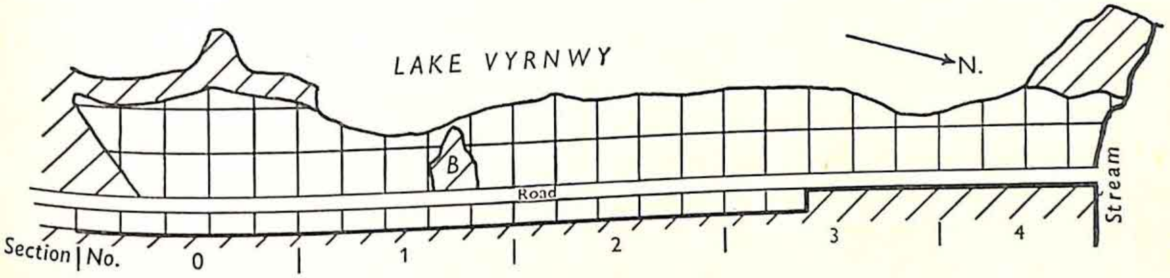


Fig. 1. Sketch-map of the trapping area with co-ordinates shown at 20 yd. intervals. *Microtus* was absent from all hatched areas, but those on the lake side of the road are scrub, suitable for *Clethrionomys*. Other hatched areas are mature woods in which *Apodemus* only is known to have occurred. The road is 15 ft. wide with 3 ft. grass verges which are included in the trapping area. B = blackthorn.

This general plan was modified in various respects, only two of which need be mentioned. (*a*) After the first trapping in March and April 1949 the whole area was retrapped in a single operation, one trap being allotted at random within areas of 20 yd × 10 yd. (*b*) In May and June 1949 we removed part of the *Microtus* population which ranged over the area south-east of the blackthorn, in sections 0 and 1 (Fig. 1).

The animals were marked by means of numbered leg rings, except in the first summer, when a system of toe-clipping or ear-marking was adopted. In going through the individual records we have found no reason for believing that any errors were caused by the loss of rings or other marks of identification, except in the case of the sample taken from the *Clethrionomys* population in September 1948.

Some idea of the sampling effort may be gathered from Table 1, which shows that the area was well covered by traps. It will be seen that the total number of traps entered, including those found sprung but unoccupied, is always less than the number of occasions on which capture might have occurred, i.e. when allowance is made for the number of times each trap was visited and reset. These facts indicate that at least one point is likely to have fallen within the range of each animal (individual *Microtus* ranged over areas of at least 10–15 yd. diameter); also that competition for traps was not so severe that the activities of one species seem likely to have introduced serious error into the sampling of another. There was a high proportion of immediate recaptures when the traps were lifted and reset

Table 1. *Summary of trapping results*

Trapping dates	Interval between trappings (units of 28 days)	No. of individuals trapped				No. of immediate re-captures	No. of traps sprung*	Total no. of traps entered	No. of points prebaited	No. of traps set	No. of visits to each trap	
		Microtus	*Clethrionomys*	*Apodemus*	Shrews						Noon to dusk	Morning
1948: 16–19 June		107	34	0	29	41	20	231	236	236	2	0
26–30 July	1·5	45	43	0	12	11	12	123	195	195	1	0
1–6 Sept.	1·3	85	64	12	13	44	32	250	326	326	0	1
13–16 Oct.	1·5	69	53	5	19	20	30	196	180	180	†	1
24–26 Nov.	1·5	67	22	3	18	0	55	165	205	205	0	1
1949: 20–23 Mar.	4·1	106	25	6	20	33	14	204	302	274	0	1
23–26 Apr.	1·2	125	24	3	13	30	23	218	305	238	0	1
20–23 May	1·0	99	21	3	17	14	23	177	220	338	0	1
16–20 June	1·0	117	14	2	23	2	31	189	415	362	1	1
20–24 July	1·1	98	29	10	66	19	23	245	224	392	1	1
29 Aug.–2 Sept.	1·5	127	46	12	53	14	24	276	246	401	1	1
16–21 Nov.	2·8	190	58	20	17	31	23	339	679	573	2	1
1950: 18–24 Apr.	5·5	26	28	14	13	18	24	123	645	368	†	1

* Traps sprung, but empty when examined. † Some traps were also visited on the day they were set: 60 in October 1948; 39 in April 1950.

Table 2. *Distribution of* Microtus *recaptures according to the interval since they were last captured* (m_{xt}). *(♂♂ and ♀♀ combined)*

Month when last captured (x)	Month of capture (t)												
	June	July	Sept.	Oct.	Nov.	Mar.	Apr.	May	June	July	Sept.	Nov.	Apr.
1948: June	—	12^3	7	4	—	1	1	—	—	—	—	—	—
July	—	—	10	—	1	—	—	—	—	—	—	—	—
Sept.	—	—	—	19	8	4	3	—	—	—	—	—	—
Oct.	—	—	—	—	11^1	9	1	2	—	—	—	—	—
Nov.	—	—	—	—	—	14	6	2	1	1^1	—	—	—
1949: Mar.	—	—	—	—	—	—	46^1	11^1	4^2	—	—	1	—
Apr.	—	—	—	—	—	—	—	34^5	18^1	—	—	—	—
May	—	—	—	—	—	—	—	—	34^{10}	3	1	—	—
June	—	—	—	—	—	—	—	—	—	40^2	5	1	—
July	—	—	—	—	—	—	—	—	—	—	56	11	—
Sept.	—	—	—	—	—	—	—	—	—	—	—	69^2	—
Nov.	—	—	—	—	—	—	—	—	—	—	—	—	19^1
Total marked (s_t)	—	12^3	17	23	20^1	28	57^1	49^6	57^{13}	44^3	62	82^2	19^1
Total unmarked (u_t)	107^{11}	33^1	68^3	46^5	47^2	78^2	68^3	50^4	60^{12}	54	65	108	7
Total catch (C_t)	107	45	85	69	67	106	125	99	117	98	127	190	26
Total released (R_t)	96	41	82	64	64	104	121	89	92	95	127	188	25

at other points on successive days in the summer of 1948 and in March and April 1949. At other times the number of these recaptures was minimized by putting the animals first captured into tins (containing a supply of food and bedding) until the traps had been reset elsewhere. In assembling the data these immediate recaptures within the period covered by any one field trip were not counted.

It will be seen from Table 1 that the majority of the animals trapped consisted of *Microtus*, followed by *Clethrionomys* and the shrews, *Sorex araneus*, *S. minutus* and *Neomys fodiens*, these five species being found at points scattered all over the area. The shrews, however, were mostly found dead in the traps and few were released alive. The wood-mouse, *Apodemus sylvaticus*, a relatively wide-ranging species, occurred in small numbers as an invader from its natural habitat outside the area. The only data suitable for analysis were thus obtained from the *Microtus* and *Clethrionomys* populations.

3. Statistical analysis

The first trapping took place on 11–14 May 1948 and the last on 18–24 April 1950, but since section 4 of the area was omitted on the first occasion, the origin of the sampling chain was taken for the purpose of this analysis at 16–19 June 1948. The samples were taken at unequal intervals of time, and although this in no way affects the methods of estimation which will be employed, it is necessary to have some approximate and simple time scale, so that a parameter such as the survival factor, determined for a particular interval, may be expressed in some unit of time for comparative purposes. Working in units of 28 days from an origin at 16–19 June 1948, the scale in the second column of Table 1 was found to correspond satisfactorily with the true dates at which the samples were taken. In addition, a purely arbitrary scale is a great convenience in order to simplify the equations for estimating the various parameters and the actual computational procedure. Thus, commencing in June 1948, the trappings were carried out at $t_0, t_1, t_2, \ldots, t_{12}$, and we will label the successive samples with these suffixes, so that the June sample is regarded as having been taken at $t = 0$, the July sample at $t = 1$, and so forth, the chain ending with the sample taken in April 1950, at $t = 12$.

Because a number of animals died in the traps or were purposely removed from the area, it is necessary to adapt some of the equations and methods given in the previous two papers in order to allow for these losses.* The corrections which must be made are, however, relatively simple, and the same symbolism with slight modifications can be used as in the theoretical development given earlier. Thus, the results of each trapping will be based on the numbers of animals captured, either alive or dead, when the traps came to be examined.

Let C_t = the number of individuals captured at time t,

d_t = the number dead in the traps or removed from the population,

and R_t = the number released alive, so that $C_t - d_t = R_t$.

The C_t individuals captured at time t can be divided into the number u_t which had not been previously captured, and the number s_t which had been captured at least once. Throughout the present work, except in one instance, these recaptures were grouped according to the interval of time since they were *last* captured (Method B: I, § 4). Thus, if

---

* In all the tables given in this paper, these losses are indicated by a superscript to the actual number caught. Thus, an entry such as 30^4 is to be read as: 30 animals were captured of which 4 were either found dead in the traps or removed from the population.

m_{xt} = the number of animals captured at time t which were last captured at time x $(x = 0, 1, 2, \ldots, t-1)$,

we have
$$\sum_{x=0}^{t-1} m_{xt} = s_t \quad \text{and} \quad s_t + u_t = C_t.$$

We will also define

ψ_t = the total number of animals captured at least once which are alive in the population as a whole at time t,

and putting
$$y_t = u_t - d_t,$$

then, immediately after the release of the R_t living individuals at time t, the total number of this class in the population will be $\psi_t + y_t$. Then, if

P_t = the survival factor over the interval of time t to $t+1$,

we expect $P_t(\psi_t + y_t)$ of these to be alive at the time of the next trapping. Similarly, if

N_t = the total number of individuals alive in the population at the trapping at time t,

$N_t - d_t$ will be alive at the time of release, of which $P_t(N_t - d_t)$ will be alive at time $t+1$. This is, of course, to adopt as before a purely deterministic model for the population. Lastly, we put

B_t = the dilution factor over the interval t to $t+1$, so that the number of new entries alive at time $t+1$ is $B_t(N_t - d_t)$.

Thus
$$N_{t+1} = (P_t + B_t)(N_t - d_t)$$
and
$$\lambda_t = N_{t+1}/(N_t - d_t) = P_t + B_t.$$

(A) *The* Microtus *population*

The first important point to be decided was whether there was any evidence in these data to support the basic assumption which is made in the theoretical development, that marked and unmarked animals were captured with equal facility. It has been shown by Chitty (1952) that in the case of a particular *Microtus* population the unmarked animals were deficient in the samples taken during the autumn and winter of 1937–8. The effect of this non-random sampling was that during a period of the year when no breeding was taking place, it nevertheless appeared as though dilution through new entries was still occurring. This phenomenon was attributed to a greater degree of relative trap-shyness that persisted among the unmarked individuals until February. The methods of trapping, however, which were used at that time were not so developed as they later became, and it was hoped that the system of prebaiting adopted in the present series would have largely overcome any degree of trap-shyness.

Before this point is examined, the biology of the *Microtus* population must be considered. The breeding season of 1948 had ended by the beginning of September. It seems likely, however, that a certain number of young born towards the end of the season might still have been growing up to enter the trappable population between September and October. Breeding started in April 1949, and a few young had grown up to be trapped on 20–23 May. No young had been caught in April. Thus, between October 1948 and April 1949, the

Microtus population should have been decreasing in numbers merely owing to the death-rate, and any dilution factors which can be estimated during this period should therefore be zero.

(i) *Preliminary analysis*

A preliminary analysis of the combined male and female population was therefore first carried out by a method of estimation which is very similar to that described in II, §5*b*. We there considered a table of m_{xt} values when the data for a long chain of samples are grouped by Method B (I, §4), and for the purposes of rough estimation we utilized only the last two entries in each column of the table, and determined the successive values of P_t and N_t from a series of overlapping triangles of m_{xt} entries. Although this method is satisfactory when relatively large numbers of animals are recaptured in the various classes, a slight difficulty was met with in applying it to some of the data analysed here, owing to the fact that in some tables certain of the m_{xt} entries for $x = t-2$ were zero, while one or more of the entries in the same column for $x < t-2$ still remained > 0 (e.g. the sample for October 1948 in Table 2). Under these conditions unsatisfactory estimates of N_t and P_t will be obtained, and a modified method of estimation was therefore adopted, which is described in an appendix to this paper.

In principle this method is similar to that described in II, §5*b*, but at the same time it utilizes more of the information contained in the original table. From the latter we form a new table by retaining each $m_{t-1,t}$ entry and then summing each column of m_{xt} over the values of $x = 0$ to $x = t-2$. Thus, for each value of $t > 1$, we have the pair of observed values

$$m_{t-1,t}, \quad n_t = \sum_{x=0}^{t-2} m_{xt} \quad (n_t + m_{t-1,t} = s_t).$$

Then from these entries and the number of animals captured (C_t) and released (R_t), together with the value of $y_t = u_t - d_t = R_t - s_t$, we first of all calculate the maximum likelihood estimate

$$\hat{\psi}_t = \frac{n_{t+1} R_t}{m_{t,t+1}} + s_t \quad (t = 0, 1, 2, \ldots, T-1).$$

Then, the successive survival factors are given by

$$\hat{P}_t = \hat{\psi}_{t+1} / (\hat{\psi}_t + y_t) \quad (t = 0, 1, 2, \ldots, T-2),$$

and

$$\hat{N}_t = \hat{\psi}_t C_t / s_t \quad (t = 1, 2, 3, \ldots, T-1).$$

Instead of this estimate of N_t, which is known to be positively biased (II, §1), we may also use the type of adjusted estimate

$$\check{N}_t = \hat{\psi}_t (C_t + 1)/(s_t + 1),$$

though we have not done so in the rough, preliminary analysis of the present set of data. Once these estimates of P_t and N_t have been calculated, which may be done very quickly, the values of λ_t follow from

$$\lambda_t = N_{t+1} / (N_t - d_t),$$

and hence those of $B_t = \lambda_t - P_t$ by subtraction. In a preliminary analysis of this nature it will probably not be necessary to calculate the sampling variances of these estimates of P_t and N_t, though the expressions for doing this are also given in the appendix.

The results of applying these methods to the data in Table 2 are presented in Table 3, where the estimates of ψ_t, N_t, λ_t, P_t and B_t are entered opposite the months of the year in which the respective samples were taken. In reading this table, as well as others given later in this paper, it is to be noted that the estimates λ_t, P_t and B_t refer to the intervals between the samples, whereas ψ_t and N_t refer to the state of the population at the time the actual sample was taken. Thus, in Table 3, $\hat{P}_0 = 0{\cdot}424$ is the estimated survival factor over the interval between the June and July samplings, and similarly $B_1 = 0{\cdot}661$ is the dilution factor between those of July and September. Although, strictly speaking, $\hat{\psi}_t$ and $\hat{N}_t$ can only take integral values, it is convenient to retain at least one decimal place in order to distinguish figures which are estimates from those which are known, true values.

Table 3. *Estimates of population parameters for* Microtus *from data in Table* 2

Month	t	$\hat{\psi}_t$	$\hat{N}_t$	λ_t	$\hat{P}_t$	B_t
1948: June	0	—	—	—	0·424	—
July	1	40·70	152·6	1·153	0·492	0·661
Sept.	2	34·26	171·3	1·343	0·759	0·584
Oct.	3	75·36	226·1	1·273	0·722	0·551
Nov.	4	84·00	281·4	0·719	0·413	0·306
1949: Mar.	5	52·87	200·2	1·221	0·856	0·365
Apr.	6	110·38	242·1	0·927	0·626	0·301
May	7	109·21	220·6	0·646	0·444	0·202
June	8	66·30	136·1	1·077	0·530	0·547
July	9	54·18	120·7	1·495	0·817	0·678
Sept.	10	85·93	176·0	1·080	0·543	0·537
Nov.	11	82·00	190·0	—	—	—

For our present purposes, it is the estimates of B_t given in Table 3 which are important. It will be seen that the individual figures remain positive throughout the chain, whereas we should have expected to obtain, between October 1948 and April 1949, estimates of B_3, B_4 and B_5 which, on the average, were zero. That these estimates remain consistently positive is sufficient to suggest that our assumed random sampling of both marked and unmarked animals in the population was not in fact fulfilled, and it is therefore necessary to examine the results for the winter months in greater detail.

A method of testing for the absence of dilution in the population sampled has been given in a previous paper (II, § 5*a*). The method was there described in terms of a population which was decreasing over the whole sampling chain through the operation of a variable death-emigration rate, this rate being assumed to fall equally on all subclasses of marked and unmarked individuals in the population. It was also assumed that no deaths were caused by the method of sampling. In the present example there are a few accidental deaths, and the hypothesis that there was no dilution applies to a period in the middle of a chain of samples. These complications cause little difficulty, however, provided that the number of accidental deaths is small, and that we tabulate the number of recaptures of all individuals marked before the date at which the population is assumed to decrease.

We will assume that no dilution was occurring in the *Microtus* population from October 1948 until the next summer, and that all living individuals in the population were exposed equally to the risk of capture. In the sample taken in May 1949, it was possible to distinguish without any difficulty the young of the year, born during April and early May, from the adults which had formed part of the overwintering population. The May sample was therefore considered only in terms of these adults. The recaptures in the samples for October, November, March, April and May were then grouped according to the time they were *first* captured and marked, and males and females were treated separately. The data for the two sexes are given in Table 4.

Table 4. *Distribution of* Microtus *recaptures October* 1948–*May* 1949 *according to the month they were first marked and released* (*adult animals only in May*)

Month when first marked	Month of capture					Total recaptures
	Oct.	Nov.	March	April	May	
	Males					
June–September	14	8	7	4	2	35
October	—	5^1	4	2	1	12
November	—	—	3	2	4	9
March	—	—	—	14^1	7	21
April	—	—	—	—	12	12
No. unmarked	25^4	29^1	38^1	34^2	21	
Total catch	39^4	42^2	52^1	56^3	47	
	Females					
June–September	9	6	5	8	1	29
October	—	1	5	3	3	12
November	—	—	4	7	3	14
March	—	—	—	17	7	24
April	—	—	—	—	9	9
No. unmarked	21^1	18^1	40^1	34^1	16	
Total catch	30^1	25^1	54^1	69^1	39	

Now, if our assumptions are correct for the type of population we are dealing with, we should expect that the number of marked animals recaptured from each batch of releases would form a constant proportion of the total catch in each sample (II, § 5*a*). Thus, neglecting any effects from the small number of accidental deaths, we may calculate the expected number of recaptures for each cell of the table from the mean proportion recaptured in each row. For example, a total of 35 males, originally marked between June and September, was recaptured in the five samples consisting of a total of 236 animals, a mean proportion of 0·1483. The 'expected' number of this class caught in October is therefore $0{\cdot}1483 \times 39 = 5{\cdot}78$, compared with the observed number of 14. In a similar fashion, the remaining expected numbers of recaptures can be calculated, the expected number of unmarked animals being obtained finally by subtraction. The observed and expected frequencies may then be used to calculate χ^2 in the usual way, the number of degrees of freedom being $\frac{1}{2}c(c-1)$ for a table consisting of c columns. Since, however, the expected numbers for the individual cells of the present tables are small, we have compared the totals for marked and unmarked animals. The following were the results for the two sexes, the expected numbers in each case being given in brackets after the observed frequencies:

Sample taken in	Males			Females		
	Marked	Unmarked	Total	Marked	Unmarked	Total
1948: Oct.	14 (5·78)	25 (33·22)	39	9 (4·01)	21 (25·99)	30
Nov.	13 (8·79)	29 (33·21)	42	7 (4·95)	18 (20·05)	25
1949: Mar.	14 (13·90)	38 (38·10)	52	14 (15·35)	40 (38·65)	54
Apr.	22 (26·38)	34 (29·62)	56	35 (34·95)	34 (34·05)	69
May	26 (34·14)	21 (12·86)	47	23 (28·75)	16 (10·25)	39

For the males $\chi^2 = 24{\cdot}7$, and for the females $\chi^2 = 12{\cdot}8$; and since there are 4 degrees of freedom in each case, $P < 0{\cdot}001$ and $< 0{\cdot}02$, $> 0{\cdot}01$, respectively. Overall, therefore, there is a significant departure from expectation, and in both sexes the deviations show that in the earlier samples more marked animals were being captured than expected.

These results, together with the non-zero values of the winter dilution factors B_t which have been calculated earlier, show that an apparent dilution of the *Microtus* population had been taking place during the winter of 1948–9, at a time of the year when no new individuals should have been entering the population at risk of capture. Judging from our previous experience of trapping this species, the most probable explanation of this phenomenon is that marked and unmarked animals were not being caught with equal facility, though the reasons for this difference in behaviour are at present obscure. For our present purposes, however, the most important conclusion to be drawn from the presence of this apparent dilution is that the assumptions made in the mathematical model are not valid for this population. If this phenomenon was present during the winter months, we cannot exclude the possibility that it may also have been present during the breeding season, when true dilution of the population must have been occurring. However, there appears to be no obvious way of testing whether or not this was the case. In these circumstances we must abandon any idea of estimating the total number of *Microtus* on the area by any of the methods which have been previously given, and which are based on the assumption of a purely random sampling of all classes in the population.

(ii) *Data for the marked individuals*

The results so far obtained from the analysis of these data for *Microtus* raise a further important question. If the chances of capturing a marked or an unmarked individual are not the same, is this difference necessarily confined to these two broad categories? We can conceive of the interaction between a population of animals and a population of traps in terms of some kind of 'learning curve'. The individuals might become more and more accustomed to the presence of these unnatural objects and to the actual experience of being trapped. The chance of capturing a marked animal might therefore depend on the number of times it had previously been trapped. If this were so, then it would be impossible to apply any methods of estimation based on the assumption that the individuals were being randomly sampled like counters in a drum.

The first way in which we attempted to answer this question was by considering separately all those individuals which were captured, marked for the first time, and released alive between September 1948 and May 1949, only adults being included in this last month. This group of releases was chosen merely because it seemed likely that they would have a more homogeneous age constitution than that of groups released during the breeding season. The animals falling into this group, which were released at each of the trappings in October, November, March, April and May, were then tabulated according to the number of times they had been trapped. Since two trappings were carried out in both March and April, an individual animal could have been released in, for instance, March, having been trapped twice in that month, apart from its previous history.

Then, if a total of r_t animals are released at time t, of which a_i have been trapped i times ($i = 1, 2, 3, \ldots; \sum_i a_i = r_t$), and if a total of c_{t+1} of these are recaptured at time $t+1$, we should expect, under a hypothesis of a purely random sampling, to find that $a_i c_{t+1}/r_t$ would be recovered out of the a_i originally released, provided that there was no differential mortality between the various classes. The numbers of individuals released each month are given in Table 5, together with the number of these recaptured at the next trapping of the area, the expected numbers of the latter, based on the distribution at the time of release, being given in brackets. Comparing the observed and expected frequencies, it can be clearly seen that there was no obvious tendency for animals which had been trapped a number of times to be caught more easily at the next retrapping. It is to be noted, however, that the minimum interval between the trappings is 28 days, and it does not necessarily follow that we should have obtained the same result if the intervals had been very much shorter.

Table 5. *The distribution of* Microtus *released according to the number of times they had been trapped, and the number of these recaptured at the next trapping of the area.* (♂♂ *and* ♀♀ *combined: adult animals only in May*)

No. of times trapped	Released Oct.	Caught Nov.	Released Nov.	Caught Mar.	Released Mar.	Caught Apr.	Released Apr.	Caught May	Released May	Caught June
1	41	6 (6·1)	45	7 (8·4)	63	24 (28·6)	57	20 (16·3)	35	14 (13·3)
2	13	2 (1·9)	12	3 (2·2)	25	16 (11·4)	33	6 (9·4)	29	13 (11·1)
3	—	—	2	1 (0·4)	9	5 (4·1)	16	5 (4·6)	5	1 (1·9)
4	—	—	—	—	1	0 (0·4)	11	3 (3·1)	5	1 (1·9)
5	—	—	—	—	1	0 (0·4)	2	0 (0·6)	2	0 (0·8)
Total	54	8 (8·0)	59	11 (11·0)	99	45 (44·9)	119	34 (34·0)	76	29 (29·0)

Expected number of recaptures given in brackets.

This evidence, which is satisfactory so far as it goes, is hardly sufficient, since we are utilizing only a portion of the data over a limited period of time. There is another method of approach, which is of much greater interest. Suppose we neglect the number of animals which were captured unmarked, namely, the u_t figures in Table 2, and confine our attention to the marked animals which were captured at each of twelve trappings between July 1948 and April 1950, inclusive. These represent a chain of twelve random samples from the total marked population on the area, and we will now regard this subpopulation as being the only one with which we are concerned. Then, from the information yielded by these samples, we can, in theory, estimate the total number of individuals in this population, and the survival and dilution factors within it for each interval of time. With reference to this subpopulation we can use exactly the same symbolism as before, the total number of individuals at time t being N_t from which a sample of C_t has been withdrawn. Similarly, in subdividing the sample of C_t into the various categories, all those animals which had been caught *only once previously* are regarded as new members (corresponding to the unmarked (u_t) class in the population as a whole), while all those which had been captured *at least twice previously* are regarded as old members (corresponding to the marked class (s_t) in the population as a whole). These old members can then be separated into the various m_{xt} classes, which now represent the re-recaptures at time t, which were last recaptured at time x. Once the principle has been grasped, the methods of assembling the data are simple. The occasion on which an individual animal was first ringed is no longer of any interest and only re-recaptures need be considered. By disregarding all animals which were not recaptured at least once, the amount of data is, of course, seriously reduced, and in the present instance we have therefore pooled the data for males and females, the resulting distributions being given in Table 6.

Table 6. *Distribution of* Microtus *re-recaptures according to the interval since they were last recaptured* (m_{xt}). *Marked population only; ♂♂ and ♀♀ combined*

Month when last captured (x)	Month of capture (t)											
	July	Sept.	Oct.	Nov.	Mar.	Apr.	May	June	July	Sept.	Nov.	Apr.
1948: July	—	3	—	—	—	—	—	—	—	—	—	—
Sept.	—	—	6	1	—	—	—	—	—	—	—	—
Oct.	—	—	—	5	2	—	—	—	—	—	—	—
Nov.	—	—	—	—	7	2	—	1	—	—	—	—
1949: Mar.	—	—	—	—	—	15	3^1	1	—	—	1	—
Apr.	—	—	—	—	—	—	13^2	9	—	—	—	—
May	—	—	—	—	—	—	—	16^5	2	—	—	—
June	—	—	—	—	—	—	—	—	22^2	—	1	—
July	—	—	—	—	—	—	—	—	—	23	4	—
Sept.	—	—	—	—	—	—	—	—	—	—	29^1	—
Nov.	—	—	—	—	—	—	—	—	—	—	—	9
s_t	—	3	6	6	9	17	16^3	27^5	24^2	23	35^1	9
u_t	12^3	14	17	14^1	19	40^1	33^3	30^8	20^1	39	47^1	10^1
Total catch (C_t)	12	17	23	20	28	57	49	57	44	62	82	19
Total released (R_t)	9	17	23	19	28	56	43	44	41	62	80	18

Among the various parameters which we may estimate from these data for the marked animals there is one, closely connected with the dilution factor B_t for this subpopulation, which is of particular interest, since the correct answer is known. This parameter we will define as

Z = the number of new additions to the marked population which were made at the sampling at time t.

Now we know the actual number of animals which were marked for the first time and released at each trapping, and we can therefore compare the estimates of Z_t, which are made from the distribution of the re-recaptures, with the true values. Any marked discrepancy between these figures, when compared with the standard errors of the estimates, would indicate that the sampling of the marked population was not at random, in which case no valid estimates of any population parameter for *Microtus* could be made from this particular set of capture-recapture data.

The estimates of Z_t may be very quickly obtained from the m_{xt} values given in Table 6. As a first step, we obtain, as before,

$$n_t = \sum_{x=0}^{t-2} m_{xt} \quad (n_t + m_{t-1,t} = s_t).$$

Then, from the total number of animals captured in the sample (C_t), the number released (R_t), and the number of accidental deaths or removals (d_t), together with the additional figure $y_t = R_t - s_t$, we calculate first of all

$$\hat{\psi}_t = \frac{n_{t+1} R_t}{m_{t,t+1}} + s_t,$$

$$p_t = s_t / C_t, \quad q_t = 1 - p_t,$$

and from these, as shown in the appendix to this paper,

$$\hat{Z}_t = (\hat{\psi}_t + y_t)/p_{t+1} - \hat{\psi}_t/p_t + d_t \quad (t = 1, 2, 3, \ldots, T-1),$$

and, for the variance of this estimate,

$$V(\hat{Z}_t) = \left(\frac{\hat{\psi}_t + y_t}{p_{t+1}}\right)^2 \left\{\frac{1}{s_{t+1}}\left[\left(\frac{p_t - p_{t+1}}{p_t}\right)^2 \frac{n_{t+1}}{m_{t,t+1}} + q_{t+1}\right]\right\} + \left(\frac{\hat{\psi}_t}{p_t}\right)^2 \frac{q_t}{s_t}.$$

The results for the *Microtus* data are presented in Table 7, where the origin of the arbitrary time scale is now taken to be July 1948. The first estimate of Z_t which can be calculated is for September 1948, the last for November 1949. The estimated total number of *Microtus* marked for the first time and released into the population over this period is 609·4, whereas

Table 7. *Estimates, from the data in Table 6, of the number of* Microtus *which were marked for the first time and released at each trapping* ($\hat{Z}_t$)

Month	t	$\hat{\psi}_t$	$\hat{Z}_t$	True no. released	Deviation Δ	Standard error of $\hat{Z}_t$
1948: Sept.	1	3·00	48·2	65	−16·8	±13·5
Oct.	2	10·60	51·4	41	+10·4	±19·1
Nov.	3	11·42	38·9	45	− 6·1	±13·6
1949: Mar.	4	12·73	66·8	76	− 9·2	±13·4
Apr.	5	29·92	111·8	65	+46·8	±26·5
May	6	45·56	19·6	46	−26·4	±20·7
June	7	31·00	35·5	48	−12·5	± 8·4
July	8	24·00	69·5	54	+15·5	±10·6
Sept.	9	35·83	78·8	65	+13·8	±15·3
Nov.	10	35·00	88·9	108	−19·1	±23·3
Total			609·4	613		

the actual number released was 613, a remarkably good agreement. Comparing the individual $\hat{Z}_t$ with the true values, it can be seen that on the whole they follow the trend of the latter reasonably well, the worst discrepancies being in the months of April and May 1949. In regard to the errors of these estimates, there is one point to be noted. The expression given above for $V(\hat{Z}_t)$ is derived from a set of maximum-likelihood equations which are based on the assumption that the probability of obtaining the observed results at each sampling is given by the appropriate term of a multinomial distribution. In actual fact, we are drawing samples from a finite population, and in using this expression, therefore, we may tend to overestimate the standard errors of $\hat{Z}_t$, if the proportion of the population sampled each time was at all large.

Now, the true proportion sampled each time, namely, $f_t = C_t/N_t$, is unknown and can only be estimated from the data. Thus we have s_t/ψ_t as an estimate of f_t, and from this it will be seen that no estimate can be made of f_T for the last sample in the chain, upon which the estimate $\hat{Z}_{T-1}$ is in part based. Moreover, if $n_{t+1} = 0$, which, it will be seen from Table 6, occurs on three occasions in the present series, then $\psi_t = s_t$, and it would appear that the entire population at risk was sampled at time t, a conclusion which seems somewhat unlikely, though not impossible, in data of this type. The simplest way of proceeding is therefore to obtain a pooled estimate of $\bar{f}_t$ from the available ψ_t, and apply the average correction $1-\bar{f}_t$ to each of the individual variances. Thus, from Tables 6 and 7 we have $\Sigma s_t/\Sigma\psi_t = 0{\cdot}694$ and $1-\bar{f}_t = 0{\cdot}306$, from which we have the adjusted standard errors of $\hat{Z}_t$ given in Table 7. A comparison of these with the deviations of $\hat{Z}_t$ from the true values suggests that the latter are of much the order we might expect. Actually $\Sigma s^2(\hat{Z}_t) = 3007$, compared with $\Sigma\Delta^2 = 4351$, and considering the approximate nature of the correction term employed, this seems a sufficiently good agreement.* Thus, by treating the marked *Microtus* as a population from which we were drawing a series of samples, we have obtained estimates of a parameter that are sufficiently close to the known, true values for us to conclude that the behaviour of this part of the total population could be regarded as satisfactory, since there was no evidence of any marked departure from the assumption of a purely random sampling of these individuals.

The results of this analysis of the data for the *Microtus* population on this particular area can be summarized as follows:

(1) There was a difference in the behaviour of marked and unmarked animals. A portion of the population was being trapped less easily than the remainder, at any rate during the winter months, and we cannot exclude the possibility that the same phenomenon was also present during the remainder of the year. This being so, no estimates can be made of the total *Microtus* population on the area by methods of estimation which are based on the assumption of a random sampling of marked and unmarked animals.

(2) Once trapped, however, the *Microtus* population appeared to behave satisfactorily from our present point of view. The samples obtained from the marked population yielded results which were very much what we might have expected if we had been randomly sampling counters from a drum, instead of trapping living animals. The data for marked

---

* It is possible that this estimate of $\bar{f}_t = 0{\cdot}694$ is too large, owing to the occurrence of three zero values of n_{t+1} in this series. If we were to neglect the three corresponding values of ψ_t and s_t, we should have $\bar{f}_t = 104/177 = 0{\cdot}587$, and the adjusted $\Sigma s^2(\hat{Z}_t) = 4059$. The latter value of $\bar{f}_t$ agrees very well with the estimate of $\bar{f}_t = 0{\cdot}567$ which is made in the next section and which is based on the recaptures of all marked individuals. The present estimates are based on the recaptures of those individuals in the subpopulation of marked animals which were marked at least twice.

Microtus may therefore be used to estimate such parameters as the death-rate at different seasons of the year, or the expectation of further life of the individuals released at some given date. Such estimates, however, will be applicable, strictly speaking, only to the trappable portion of the *Microtus* population.

(iii) *Methods of calculating the death-rate and of constructing life-tables*

As an example of a simple method of estimating the death-rate per unit of time in the population as a whole, we may use the estimates of the survival factors P_t given in Table 3. These estimates are based essentially on the numbers in the various classes of marked individuals which were recaptured from the known number of animals released (e.g. II, §§ 1 and 5), and no use is made of the observed proportion of unmarked animals occurring in the sample of C_t taken at time t. This last point is of importance since, if this was not the case, the estimates of P_t would be just as untrustworthy as the estimates of total numbers for this particular *Microtus* population. The values of $\hat{P}_t$ given in Table 3 refer, however, to varying intervals of time, and for comparative purposes, therefore, it is necessary to express them in terms of some common unit. If we define

$\mu_t =$ the force of mortality between t and $t+dt$, which we assume to remain constant over the interval t to $t+w$, we have

$$P_t = e^{-w\mu_t}.$$

Then

$$\mu_t = -\log_e P_t/w,$$

with

$$V(\mu_t) = V(P_t)/(w^2P_t^2),$$

the expression for calculating $V(\hat{P}_t)$ being given in the appendix to this paper. These variances may, however, require correction in order to allow for the finite nature of the population sampled, particularly if a relatively large proportion of the marked population was being captured each time. This proportion can only be estimated from the data, and, as before, we have employed the estimated average proportion $\bar{f}_t$ sampled over the whole period. Thus, in Table 3, the values of $\hat{\psi}_t$ are estimates of the total number of marked *Microtus* in the population at time t. The sum of the available values of $\hat{\psi}_t$ is 795, and from Table 2 the sum of corresponding s_t is 451. Hence, $\bar{f}_t = 0{\cdot}567$, and we have corrected the estimated variances of μ_t by multiplying them by the factor 0·433. The following were the values of μ_t expressed in a unit of time of 28 days, together with their standard errors:

Month	Death-rate $\mu_t \pm s$
1948: June–	$0{\cdot}572 \pm 0{\cdot}152$
July–	$0{\cdot}545 \pm 0{\cdot}174$
Sept.–	$0{\cdot}184 \pm 0{\cdot}143$
Oct.–	$0{\cdot}217 \pm 0{\cdot}154$
Nov.–	$0{\cdot}216 \pm 0{\cdot}039$
1949: Mar.–	$0{\cdot}129 \pm 0{\cdot}089$
Apr.–	$0{\cdot}468 \pm 0{\cdot}116$
May–	$0{\cdot}812 \pm 0{\cdot}087$
June–	$0{\cdot}577 \pm 0{\cdot}056$
July–	$0{\cdot}135 \pm 0{\cdot}041$
Sept.–Nov.	$0{\cdot}218 \pm 0{\cdot}011$

The death-rate per 28 days in the marked *Microtus* population, taking both sexes together, was high during the summer of 1948. From September onwards it fell and appeared to remain approximately constant over the winter months until April 1949. Thereafter there was a rise in the rate until July, and between September and November 1949 it was again of very much the same order as in the previous year. Although no estimate of the death-rate can be made for the period subsequent to November 1949, the rate must have increased very greatly during the second winter, since the *Microtus* population was evidently reduced to very small numbers by April 1950.

The relative constancy of the death-rate during the winter of 1948–9 is of some interest, since the population is of a more homogeneous age constitution at this season of the year than at any other, and neither sex is in breeding condition. In order to see whether there was any difference between the sexes during this time, we also estimated the average death-rate between October 1948 and April 1949 by the methods of estimation which have been previously given (I, § 5B and Appendix 1). The equations given there are expressed in terms of a constant survival factor P and an equal interval between the samplings. It is, however, easy to adapt them to the case of unequal intervals, and although the equations to be solved are somewhat more complicated, the actual method of solution remains, in principle, the same.

In order to estimate the constant survival factor P per unit of time over the period October 1948 to April 1949, we make use of the number of animals released at and between these dates, and of the recaptures of marked animals in the samples taken between November 1948 and May 1949 inclusive, due allowance being made for any accidental losses or removals (I, Appendix 2), and only adult animals being considered in May. The following were the estimates of the constant survival factor P per unit of 28 days and the values of the death-rate obtained from them. No correction has been made of the standard errors of these estimates to allow for the finite nature of population sampled:

	Survival factor $P \pm s$	Death-rate $\mu \pm s$
♂♂	$0{\cdot}725 \pm 0{\cdot}0288$	$0{\cdot}322 \pm 0{\cdot}040$
♀♀	$0{\cdot}867 \pm 0{\cdot}0289$	$0{\cdot}143 \pm 0{\cdot}033$
Combined sexes	$0{\cdot}817 \pm 0{\cdot}0182$	$0{\cdot}202 \pm 0{\cdot}022$

Evidently there was quite a marked difference in the average death-rates of the two sexes. The difference between them is $0{\cdot}179 \pm 0{\cdot}052$, which would be reckoned as significant, quite apart from the fact that we are probably overestimating the individual variances. The average death-rate for the combined sexes is $\mu = 0{\cdot}202$, while the average obtained by weighting the three individual μ_t values in the previous table for October, November and March by the reciprocal of their variances is $\bar{\mu}_t = 0{\cdot}203$.

The death-rate in the population as a whole is, however, not the most satisfactory way of investigating the changing mortality experience of a species such as *Microtus*, which, owing to the intermittence of its breeding season, must suffer a series of somewhat violent and complicated changes in the age constitution of the population during the course of the year. Thus, from the figures given earlier, it is impossible to decide whether the increase in the death-rate which occurred in the summer months of 1949 was due to an increase in the rate at which the overwintering adults were dying off, or whether it was due to a high mortality among young born during the early part of the breeding season. Indeed, both of those factors might have been operating, though in different degrees. A more satisfactory way of approaching the problem would be to construct life-tables for each batch of newly ringed animals which were released into the population and, if the data permitted, to subdivide each batch into males and females, and also into young and old. A detailed subdivision of this nature would, however, require much larger figures than those in the present series. But, as an example of the methods of estimating these life-tables, we shall compare the mortality experience of the individuals released for the first time in September, October and November 1948 with that of the releases in March, April and May 1949, considering only adult animals in this last month. These animals which were trapped for the first time in the spring of 1949 must have been alive and of trappable age, certainly in October 1948, and probably also in September. They were, however, not caught then, and in view of the phenomenon of non-random sampling during the winter months to which we have already referred, it was of interest to see whether their expectation of further life, once they were caught, differed appreciably from that of the survivors of the groups which entered the traps during the previous autumn.

The various methods of estimating the number of survivors at successive intervals of time from a known number of individuals released at some given date have already been described (II, § 2). If G_d individuals form the group released at a date d, then any recaptures of this group in the samples taken from the general population can be recognized by the presence of the mark d. These recaptures can be tabulated in various ways, and in the present series of data we assembled them in the form of the frequencies m_{dxt}, the number recaptured at time t which were last captured at time x $(d<x<t)$. We then obtained estimates of the number of survivors (N_x) of the original group by the approximate methods described in II, § 2*b*. There is, however, one further complication in the present data, and that is the occurrence of accidental deaths and removals from the population at certain trappings. The simplest method of dealing with these is best illustrated by an actual example.

In March 1949, 76 *Microtus* (37♂♂, 39♀♀) were marked for the first time and released. Of these, one individual was found dead in a trap in April, and five were trapped in June and removed from the population. The life histories of these six animals out of the original 76 are therefore known, and we can immediately write down the corresponding life-table (l'_x) figures for them. Thus, six were alive at the time of the trapping in April, five in May and five in June, and none after that date. The remaining records for this group are represented by the following table (see p. 154) of m_{xt} values, in which we arbitrarily label the trappings from April onwards as being taken at $t = 0, 1, 2, \ldots$.

It will be seen that, for instance, out of the remaining 70 animals released in March (having eliminated the six accidental deaths and removals), seven were caught in June, of which two were last caught in May and four in April, the remaining individual in the sample having escaped capture since March. The last member of the group was trapped in November, and hence the last estimate which can be made of the number of survivors is that for September. Since no members of the group were caught in April 1950, the number of survivors in November is indeterminate; but we do know that at least one individual was alive then, and we shall adopt the convention both in this, and in all other such tables, that the survivorship curve ends at the last date on which any member of the original group was recaptured. The method of computing the estimates of the number of survivors at times $x = 0, 1, 2, \dots, T-1$ from such a table has already been illustrated on a numerical example in an earlier paper, the first step being to build up the successive z_t and $\Sigma k'_{xt}$ values, starting at the bottom right-hand corner. (For a definition of these symbols, and an explanation of the way in which they arise, reference should be made to II, § 2*b*.) Then

x \ t	Apr. 0	May 1	June 2	July 3	Sept. 4	Nov. 5	$\sum_t k'_{xt}$	R_x	$\sum_{x+1}^{T} R_t$	$\check{N}'_x$	l'_x	$\check{N}_x$
Apr. 0	—	6	4	—	—	—	15·86	30	25	46·3	6	52·3
May 1	—	—	2	1	—	—	5·25	11	14	26·4	5	31·4
June 2	—	—	—	3	—	—	5·25	7	7	9·0	5	14·0
July 3	—	—	—	—	2	1	3·00	4	3	4·0	0	4·0
Sept. 4	—	—	—	—	—	0	0	2	1	4·0	0	4·0
Nov. 5	—	—	—	—	—	—	—	1	—	1	0	1
Total caught and released (R_t)	30	11	7	4	2	1						
z_t	—	1·4773	1·75	1·75	1·00	1·00						

$$\check{N}_x = R_x\left(\sum_{x+1}^{T} R_t + 1\right) \Big/ \left(\sum_{t=x+1}^{T} k'_{xt} + 1\right) + l'_x.$$

Thus, from the table above, we have, as a first step,

$$\check{N}'_0 = 30 \times 26/16{\cdot}86 = 46{\cdot}3,$$

and since from the survivorship curve of the six accidental deaths and removals, $l'_0 = 6$, we finally have $\check{N}_0 = 52{\cdot}3$ as an estimate of the number of survivors out of the 76 released in March which were alive at the time of the trapping in April. Having thus obtained the series of $\check{N}_x$, ending with the number trapped at the last date on which any of the group were recaptured, the construction of the life-table follows in the customary way. Thus, when no accidental deaths and removals are present, we merely express the $\check{N}_x$ in terms of 100 or 1000 individuals alive at the original date of release. When this is not the case, as in the present example, it is necessary to calculate the series of P_x figures, after allowing for the accidental deaths and removals which take place at each ordinate x. In June, for example, 14·0 individuals were estimated to have been alive at the time of the trapping, five individuals were removed from the population, and therefore, since 4·0 were estimated to have been alive in July, the survival factor over the interval June to July is $P_2 = 4{\cdot}0/(14{\cdot}0-5) = 0{\cdot}\dot{4}$. Then, the values of the life-table function follow from $l_x = P_0 P_1 P_2 \dots P_{x-1}$.

In order to combine the results for a number of groups of releases, we may pool the estimated $\Sigma k'_{xt}$ and observed R_x and ΣR_t in each table. Thus, for example, from the separate tables for the original data, which are not presented here, we had for the trapping which took place in March 1949:

Originally released in	$\sum_t k'_{xt}$	R_x	$\sum_{x+1} R_t$	l'_x
1948: Sept.	9·25	7	19	1
Oct.	6·25	8	10	4
Nov.	5·71	5	13	0
Total	21·21	20	42	5

From which we have, from the 'total' line,

$$\check{N} = \frac{20 \times 43}{22 \cdot 21} + 5] = 43 \cdot 7,$$

as an estimate of the total number of these releases surviving in March. Proceeding in this way, we finally obtained the following life-tables (l_x) and expectations of further life, in weeks, of individuals alive at the given dates (e_x):

Month	Individuals marked for the first time and released in				Unmarked animals captured (u_t)
	Sept., Oct. and Nov.		Mar., Apr. and May		
	l_x	e_x (weeks)	l_x	e_x (weeks)	
1948: Oct.	1·0000	12·9	—	—	46
Nov.	0·4793	17·6	—	—	47
1949: Mar.	0·2231	12·0	1·0000	9·9	78
Apr.	0·2001	8·3	0·6882	8·5	68
May	0·1582	6·0	0·4781	7·3	37
June	0·0639	—	0·2081	—	—
July	0·0372	—	0·1005	—	—
Sept.	0·0149	—	0·1225	—	—
Nov.	0·0074	—	0·0184	—	—

It is evident from the figures available that the expectation of further life of the over-wintering individuals which were caught for the first time in March, April and May differed very little from that of the survivors of those originally captured in September, October and November. This being so, we might assume that their previous mortality experience in the unmarked state was probably much the same as in the case of the autumn releases. There is, of course, no way of knowing whether or not this was true. If the sampling of marked and unmarked *Microtus* had been satisfactory during the winter months, we should have had no hesitation in making this assumption; indeed, it is implicit in all applications of the capture-recapture method. But, in the present example, we cannot be certain that a difference in the behaviour of these two classes was not associated with a difference in mortality experience. If we assume, however, that there was no difference in the mortality rates of marked and unmarked animals, we may obtain a rough estimate of the size of the

Microtus population on the area in October 1948 by applying the l_x figures for the autumn releases to the number of unmarked adult animals which were captured up to May 1949. These figures are given in the last column of the table, and by dividing each of them by the appropriate l_x entry and summing, we obtain a total of just over 1000 animals alive in October. If our l_x figures were exact, which they are not, this would be an underestimate of the true figure, since a number of overwintering adults were still being captured unmarked in June, and possibly also in July, though we have been unable to separate these individuals with any certainty from the young. Moreover, in the population at October 1948, there were also the survivors of those marked at the trappings previous to this date, which we estimated in Table 3 to be around 75 individuals. Such estimates as we are indulging in here can only be approximate, but, so far as they go, they suggest that the total *Microtus* population in October 1948 was of the order of 1000 individuals, or a population density of around 250 per acre. Comparing the former figure with the estimates of N_t in Table 3, it will be seen how greatly we would have underestimated the probable *Microtus* population on the area in the autumn of 1948 if we had assumed without any further investigation that the sampling of this species was satisfactory from the theoretical point of view.

B. *The* Clethrionomys *population*

The average range covered in its lifetime by an individual *Clethrionomys* was somewhat greater than it was in *Microtus*, and the trapping area was less well defined by ecological barriers. Nevertheless, most *Clethrionomys* were retrapped within perhaps 30 yd. of their original point of capture, and, as is shown below, the population was not recruited except at the times one might have expected it to increase by breeding.

The general biology of this species appears to be very similar to that of *Microtus*. Thus, from the examination in the laboratory of dead *Clethrionomys*, trapped at the same time on other areas at Lake Vyrnwy, there was no evidence that any breeding was taking place during the autumn and winter months; whereas pregnant and lactating females were common in the samples obtained during the summer. From October to April, therefore, we can assume that the population would be decreasing merely through deaths, and that, unless immigration were occurring, no new individuals should have been entering the population at risk of capture. In one minor respect, however, the data for *Clethrionomys* have been treated differently from those for *Microtus*. In September 1948 we marked both species by means of an ear punch, and while these holes remained perfectly clear in the case of *Microtus*, we have reason to suspect that some might have healed over in the case of *Clethrionomys*. In order to avoid any possibility of error through the loss of these marks we have therefore disregarded the September sample for this species. Thus, all those animals known to have been first marked in September are regarded as unmarked when first recaptured in later months, and the September mark is similarly neglected in all other cases.

(i) *Preliminary analysis*

The preliminary analysis of the *Clethrionomys* data is thus similar to that of *Microtus*. We first of all estimate the successive dilution factors B_t for the entire series of trappings in order to see whether these approach zero in the winter months. The data and results are given in Tables 8 and 9, and from the former it will be seen that the individual catches (C_t)

of this species are not very large, but appear to follow a fairly regular annual course, being high in July and October 1948, low in the winter months, and gradually increasing during the summer of 1949 to a peak in September and November. In April 1950, the number trapped was much the same as in the preceding April, a complete contrast to the results for the *Microtus* population (Table 2).

Table 8. *Distribution of* Clethrionomys *recaptures according to the interval since they were last captured* (m_{xt}). (*♂♂ and ♀♀ combined*)

Month when last captured (x)	Month of capture (t)											
	June	July	Oct.	Nov.	Mar.	Apr.	May	June	July	Sept.	Nov.	Apr.
1948: June	—	17^3	1	1	—	—	—	—	—	—	—	—
July	—	—	11	3	—	—	—	—	—	—	—	—
Oct.	—	—	—	14	7	6	2	—	—	—	—	—
Nov.	—	—	—	—	10	—	2	—	1	—	—	—
1949: Mar.	—	—	—	—	—	15	1	1	—	—	—	—
Apr.	—	—	—	—	—	—	10^1	1	—	—	—	—
May	—	—	—	—	—	—	—	5	4	—	—	—
June	—	—	—	—	—	—	—	—	6	—	—	—
July	—	—	—	—	—	—	—	—	—	10	—	2
Sept.	—	—	—	—	—	—	—	—	—	—	19^1	1
Nov.	—	—	—	—	—	—	—	—	—	—	—	17
Total marked (s_t)	—	17^3	12	18	17	21	15^1	7	11	10	19^1	20
Total unmarked (u_t)	34^1	26^2	41^4	4	8	3	6	7	18^1	36^1	39^5	8
Total catch (C_t)	34	43	53	22	25	24	21	14	29	46	58	28
Total released (R_t)	33	38	49	22	25	24	20	14	28	45	52	28

Turning to the estimates of B_t given in Table 9, it will be seen that for the period October–April, one estimate is negative, one positive, and one not very far from zero. If we consider an imaginary population of 100 individuals alive in October 1948, and subject to these estimated $\lambda_t = (P_t + B_t)$ and P_t values, then, according to the latter, 36·1 out of the original 100 would be alive in April 1949, and from the successive λ_t the total population size would

Table 9. *Estimates of population parameters for* Clethrionomys *from data in Table* 8

Month	t	$\hat{N}_t$	λ_t	$\hat{P}_t$	B_t
1948: June	0	—	—	0·620	—
July	1	51·7	2·458	0·627	1·831
Oct.	2	114·8	0·368	0·530	−0·162
Nov.	3	40·8	0·973	0·722	0·251
1949: Mar.	4	39·7	1·000	0·943	0·057
Apr.	5	39·7	0·811	0·639	0·172
May	6	32·2	1·195	0·667	0·528
June	7	37·3	0·777	0·429	0·348
July	8	29·0	1·643	0·357	1·286
Sept.	9	46·0	1·911	0·626	1·285
Nov.	10	86·0	—	—	—

be 35·8; in other words, on the average the change in numbers is accounted for by the operation of the death-rate. (This result is in marked contrast to that for the *Microtus* population during the same period, in which only 25 out of an original 100 would be alive in April, whereas the total population would have increased to 114.)

These preliminary results for *Clethrionomys* are promising from our present point of view, and it is unfortunate that no estimates of the dilution factor could be made for the second winter during the period November 1949 to April 1950. We may, however, calculate an upper limit to B_t for this interval. For this purpose we will define a new quantity,

$$U_t = P_t^{-1} B_t(N_t - d_t).$$

This parameter may have very little real meaning or importance in the case of the general population, since it merely represents the number of new entries $B_t(N_t - d_t)$ alive at time $t+1$ which would have been alive at time t if they had been subject to the estimated survival factor P_t. For instance, if the new entries consisted of births during the interval, this clearly would be an entirely artificial figure. Mathematically, however, it is the exact equivalent of the parameter defined as Z_t in the case of the subpopulation of marked animals, and it may be estimated by a similar type of equation. Thus we have, as shown in the appendix,

$$\hat{U}_t = R_t\left[\frac{s_t C_{t+1} - n_{t+1} C_t}{m_{t,t+1} s_t} - 1\right].$$

Now, no estimate can be made of $\hat{P}_{T-1}$ for the last interval in a chain of samples, but we can obtain $\hat{N}_{T-1}$ and $\hat{U}_{T-1}$. Thus, from the data given in Tables 8 and 9, we have for November 1949,

$$\hat{N}_{T-1} - d_{T-1} = 86{\cdot}0 - 6 = 80{\cdot}0 \quad \text{and} \quad \hat{U}_{T-1} = 52\left[\frac{19 \times 28 - 3 \times 58}{17 \times 19} - 1\right] = 5{\cdot}6.$$

Hence, under the most favourable circumstances, if $P_{T-1} = 1$, we should estimate the dilution factor between November 1949 and April 1950 as $B_{T-1} = 0{\cdot}07$, a figure which is to all intents and purposes zero.

As a check on this apparent absence of dilution during the winter months, we may now consider the recaptures of marked animals during October, November, March, April and May (adults only), when these are tabulated according to the interval of time since they were first marked. Since the number of individuals was in any case small, the data for the two sexes were combined, with the results given in Table 10. Calculating the expected numbers of marked and unmarked animals as before, we have:

Month of trapping	Marked	Unmarked	Total
1948: Oct.	12 (12·79)	41 (40·21)	53
Nov.	18 (15·11)	4 (6·89)	22
1949: Mar.	17 (17·53)	8 (7·47)	25
Apr.	21 (19·50)	3 (4·50)	24
May	15 (18·06)	6 (2·94)	21

Comparing the observed numbers with the expected given in brackets, there appears to be no very marked trend in the signs of the deviations as there was in the case of the *Microtus* data. Combining the last two classes together, owing to the small expectations in the unmarked class, $\chi^2 = 2{\cdot}27$, which for 3 D.F. is a perfectly reasonable value to obtain.

Table 10. *Distribution of* Clethrionomys *recaptures October* 1948–*May* 1949 *according to the month they were first marked and released.* (*♂♂ and ♀♀ combined*)

Month when first marked	Month of capture					Total recaptures
	Oct.	Nov.	Mar.	Apr.	May	
June–July	12	8	4	6	5	35
Oct.	—	10	12	11	8	41
Nov.	—	—	1	0	0	1
Mar.	—	—	—	4	1	5
Apr.	—	—	—	—	1	1
No. unmarked	41^4	4	8	3	6	
Total catch	53^4	22	25	24	21	

Thus, from this preliminary analysis, we may conclude that the sampling of the *Clethrionomys* population was satisfactory, and that there was no evidence of any pseudo-dilution during the non-breeding season as in the case of *Microtus*. It is perhaps worth adding that we also carried out an analysis of these data in which we assumed that no September marks had been lost, and reached precisely the same conclusions.

(ii) *Data for the marked individuals*

It is useful to have a further check on the applicability of the methods of estimation to these particular data by estimating the successive values of Z_t from the re-recaptures of the marked population. The observed data and resulting estimates are given in Tables 11 and 12,

Table 11. *Distribution of* Clethrionomys *re-recaptures according to the interval since they were last recaptured* (m_{xt}). (*Marked population only; ♂♂ and ♀♀ combined*)

Month when last captured (x)	Month of capture (t)										
	July	Oct.	Nov.	Mar.	Apr.	May	June	July	Sept.	Nov.	Apr.
1948: July	—	5	—	—	—	—	—	—	—	—	—
Oct.	—	—	4	—	2	1	—	—	—	—	—
Nov.	—	—	—	9	—	2	—	1	—	—	—
1949: Mar.	—	—	—	—	11	1	1	—	—	—	—
Apr.	—	—	—	—	—	9^1	1	—	—	—	—
May	—	—	—	—	—	—	4	3	—	—	—
June	—	—	—	—	—	—	—	4	—	—	—
July	—	—	—	—	—	—	—	—	6	—	1
Sept.	—	—	—	—	—	—	—	—	—	3	—
Nov.	—	—	—	—	—	—	—	—	—	—	7
s_t	—	5	4	9	13	13^1	6	8	6	3	8
u_t	17^3	7	14	8	8	2	1	3	4	16^1	12
Total catch (C_t)	17	12	18	17	21	15	7	11	10	19	20
Total released (R_t)	14	12	18	17	21	14	7	11	10	18	20

and it will be seen from the latter that between October 1948 and November 1949 a total of 151 bank voles were marked for the first time and released, whereas the estimated total is 150·8. The individual $\hat{Z}_t$ follow the general trend of the true values very closely, considering the small size of the samples on which they are based. In regard to the standard errors of these estimates, the same procedure was followed as in the case of the *Microtus* data. The sum of the available ψ_t is 96, and the sum of the corresponding $s_t = 67$. Hence $\bar{f}_t = 0{\cdot}698$, and the variances of $\hat{Z}_t$ were then multiplied by the factor 0·302 to give the individual standard errors in the last column of Table 12. Comparing these with the deviations of the estimates from the true values, it seems that on the whole the latter are of much the order we might expect from errors of sampling. The chief exceptions are the estimates for November 1948 and July 1949, for which the deviations are roughly twice and three times the respective standard errors. An unknown margin of error in the latter must, however, be allowed for, owing to the approximate nature of the correction employed. Overall, the sum of the squares of the deviations from the true values is 921, while $\Sigma s^2(\hat{Z}_t) = 810$. Generally speaking, therefore, there is no evidence of any marked discrepancy in these results, and if this had been a problem of estimating an unknown parameter Z_t, given only the capture-recapture data in Table 11, we should not have been led into any very serious error either as to the magnitude of this parameter or the changes it underwent during the period of observation.

Table 12. *Estimates, from the data in Table* 11, *of the number of* Clethrionomys *which were marked for the first time and released at each trapping* (Z_t)

Month	$\hat{\psi}_t$	$\hat{Z}_t$	True no. released	Deviation Δ	Standard error of $\hat{Z}_t$
1948: Oct.	5·00	42·0	37	+ 5·0	± 13·3
Nov.	4·00	16·0	4	+ 12·0	± 6·1
1949: Mar.	12·09	9·6	8	+ 1·6	± 4·2
Apr.	22·33	− 1·1	3	− 4·1	± 4·2
May	20·00	2·4	6	− 3·6	± 2·4
June	13·00	4·1	7	− 2·9	± 2·4
July	8·00	7·3	17	− 9·7	± 2·8
Sept.	6·00	53·3	35	+ 18·3	± 18·5
Nov.	5·57	17·2	34	− 16·8	± 14·1
Total		150·8	151		

(iii) *Estimation of total numbers*

The results of all the tests which we have been able to apply to these data suggest that the sampling of this species was satisfactory from the theoretical point of view. There was no evidence of any differential sampling of marked and unmarked animals, and the estimates of the numbers marked for the first time and released at each trapping were in good agreement with the true values. We may therefore apply to these data, with some degree of confidence, the methods which have previously been described for estimating the total number of individuals in the population.

Since we are here dealing with a relatively long chain of samples taken over a period of very nearly two years, during which time both the total numbers and the death-rate may have been varying quite appreciably, the appropriate methods of estimation are in principle similar to those which were briefly discussed in II, § 4. Thus, we may take each group of G_x animals which were caught unmarked and released marked at time x ($x = 0, 1, 2, 3, \ldots$) and estimate the number of these surviving at time t ($t = x+1, x+2, x+3, \ldots$) by one or another of the methods given in II, § 2. Then, if G_{xt} is the estimated number which were surviving at time t, with variance $V(G_{xt})$, the total number of marked individuals in the population at time t is clearly

$$\psi_t = \sum_{x=0}^{t-1} G_{xt},$$

and

$$V(\psi_t) = \sum_{x=0}^{t-1} V(G_{xt}),$$

since the separate groups of animals are independent. We may then calculate the total number of individuals in the population by means of the maximum-likelihood estimate

$$\hat{N}_t = \psi_t C_t / s_t \quad (t = 1, 2, 3, \ldots, T-1);$$

or, preferably perhaps, by the adjusted estimate

$$\check{N}_t = \psi_t (C_t + 1)/(s_t + 1),$$

with variance

$$V(\check{N}_t) = \check{N}_t^2 \left[\frac{u_t}{(s_t+2)(C_t+1)} + \frac{V(\psi_t)}{\psi_t^2} \right].$$

In order to carry out satisfactorily all the various steps in this computation, however, it would be necessary to have a very much larger body of data than in the present example. The number of *Clethrionomys* which were ringed for the first time and released at each sampling was small in a number of cases, as may be seen from the unmarked (u_t) figures given in Table 8. We therefore obtained a pooled estimate of the number of marked animals alive in the population in a similar way to that described for the construction of the life-tables for *Microtus*.

Each group of releases was considered separately, and having first of all found the life-table l'_x values for any individuals in the group which were accidentally killed, the recaptures of the remaining individuals were tabulated according to the interval of time since they were last recaptured. The values of $\sum_t k'_{xt}$, R_x and $\sum_{x+1}^{T} R_t$ were then computed from these entries, as illustrated above in § A (iii), by the methods given in II, § 2*b*. For a particular month, all the available values of these three quantities in the separate tables were then pooled, as well as any l'_x figures. Then, an estimate of the total number of marked animals in the population at time t is given by

$$\check{\psi}_t = SR_x \left(S \sum_{x+1}^{T} R_t + 1 \right) \Big/ \left(S \sum_{t=x+1}^{T} k'_{xt} + 1 \right) + Sl'_x,$$

the variance of this estimate (II, § 2*b*) being

$$V(\check{\psi}_t) = \psi_t^2 \left[\frac{S\Sigma R_t - S\Sigma k'_{xt}}{(S\Sigma R_t + 1)(S\Sigma k'_{xt} + 2)} \right],$$

where in both expressions the symbol S is used to indicate the pooled values. The estimates of $\check{N}_t$ and $V(\check{N}_t)$ then follow from the equations given above. The following results were obtained:

Month	$S\Sigma k'_{xt}$	SR_x	$S\Sigma R_t$	Sl'_x	$\check{\psi}_t \pm s$	s_t	C_t	$\check{N}_t \pm s$
1948: July	9·17	14	12	3	20·9 ± 1·5	17	43	51·1 ± 6·9
Oct.	16·67	12	27	0	19·0 ± 1·6	12	53	78·9 ± 12·9
Nov.	30·36	18	60	1	36·0 ± 2·6	18	22	43·6 ± 4·1
1949: Mar.	24·51	17	44	1	31·0 ± 2·3	17	25	44·8 ± 4·9
Apr.	19·10	20	34	1	35·8 ± 3·0	21	24	40·7 ± 3·9
May	14·00	13	21	2	21·1 ± 1·6	15	21	29·0 ± 3·3
June	7·60	7	17	2	16·7 ± 2·1	7	14	31·3 ± 6·2
July	8·00	11	8	0	11·0 ± 0	11	29	27·5 ± 3·6
Sept.	4·00	10	6	0	14·0 ± 1·8	10	46	59·8 ± 12·0
Nov.	7·00	18	10	1	25·8 ± 2·6	19	58	76·1 ± 11·4

One minor point is to be noted in regard to the estimates of ψ_t. At the trapping in July 1949, we have $\check{\psi}_t = 11\cdot0 \pm 0$; in other words, it seems that the entire marked population at risk was captured. It is always possible, of course, that in the case of a small population and a high degree of sampling effort, this may happen, but if so we must have $\hat{N}_t = C_t$, when we use the maximum-likelihood estimate of N_t, and in using the adjusted estimate we will have a value of $\check{N}_t < C_t$, as will be seen in the last column of the table.

The standard errors of $\check{\psi}_t$ and $\check{N}_t$ given in the table are adjusted to allow for the proportion of the population which on the average was being sampled. Thus, by summing the $\check{\psi}_t$ and s_t columns in the table, we have $\bar{f}_t = 147/231 = 0\cdot636$, and $1-\bar{f}_t = 0\cdot364$, this factor being then used to correct the variances. (It is interesting to compare the value of this factor with that used to correct the variances of the estimates of Z_t given earlier in this section. There, $1-\bar{f}_t = 0\cdot302$, this figure being based on the estimated ψ_t and observed s_t in Tables 11 and 12. Those ψ_t and s_t are not, however, the same as the ones in the present table, since they represented the individuals in the subpopulation of marked animals which were marked at least twice.) It will be seen from the estimates $\check{N}_t$ of total numbers that the trappable *Clethrionomys* population on the area was not very large, amounting to probably not much more than 100 individuals of both sexes at the peak period of numbers during the late summer and autumn. From these figures we might say that this was a population which was on the average relatively stationary in numbers, and which was merely oscillating during the two years owing to the effect of an intermittent breeding season.

In order to complete the description of this population, we also require estimates of the death-rate and dilution-rate expressed per unit of time. The successive survival factors P_t may be obtained, as before, by means of

$$P_t = \check{\psi}_{t+1}/(\check{\psi}_t + y_t),$$

and hence the death-rate,

$$\mu_t = -\log_e P_t/w_t,$$

where w_t is the true interval of time between the particular samplings to which P_t refers. We also have

$$\lambda_t = \check{N}_{t+1}/(\check{N}_t - d_t),$$

and we will define the rate of increase during the interval by

$$\rho_t = \log_e \lambda_t/w_t.$$

This rate of increase ρ_t between t and $t+dt$, which we assume to remain constant over the interval t to $t+w$, is composed of a positive quantity β_t, representing the rate of dilution, and μ_t the death-rate, so that $\beta_t - \mu_t = \rho_t$. Having calculated ρ_t and μ_t, we thus can obtain β_t very simply by subtraction. The sampling errors of these estimates, however, are more difficult and we have been unable to solve the problems involved. The difficulty arises over the question as to the degree of correlation which may exist between the successive estimates of ψ_t and N_t which are obtained from the data treated in this way. Any such correlation which exists is likely to be negative in sign, and, if we were to neglect this and merely write down the expressions for the variances of ρ_t and λ_t according to large-scale sample theory, we run the risk of underestimating the true sampling variances of the resulting β_t and μ_t. Leaving aside this question, however, the following values of ρ_t, β_t and μ_t, expressed per unit of 28 days, were obtained for the *Clethrionomys* population:

	Dilution-rate (β_t)	Death-rate (μ_t)	Rate of increase (ρ_t)
1948: June–	—	0·305	—
July–	0·475	0·283	0·192
Oct.–	−0·067	0·294	−0·361
Nov.–	0·069	0·062	0·007
1949: Mar.–	−0·009	0·071	−0·080
Apr.–	0·271	0·609	−0·338
May–	0·559	0·446	0·113
June–	0·681	0·798	−0·117
July–	1·005	0·462	0·543
Sept.–Nov.	0·320	0·229	0·091

The first point of interest to be noted is that the three dilution-rates between October and April are each essentially zero, a result which might have been expected from the preliminary analysis, but which nevertheless serves as a useful check on our original approximate estimates of B_t in Table 9, since we are now utilizing much more of the information contained in the data. After April the dilution-rate gradually increases during the breeding season of 1949. The death-rate in June and July appears to differ quite markedly in the two years, being about twice as great in 1949 as in 1948. It is interesting to find that the death-rate falls to a relatively low level during the non-breeding season between November 1948 and April 1949, in a similar way to that of *Microtus*. It then increases markedly after April, but the dilution-rates are at that time sufficient to keep the population relatively stationary in numbers until July. There is then a sharp rise in the rate of increase between July and September, which is presumably due to an influx of the young born during the latter part of the breeding season, which have grown up to become part of the trappable population. The broad pattern of these results thus confirms our previous knowledge of the biology of this species, and in addition we have obtained a numerical measure of the changes which occurred in the various population parameters during the period of observation.

4. Discussion of results

In the first two papers of the present series we were concerned with the theory of estimating population parameters from a series of samples, granted certain premises. We could also show that under conditions when these assumptions were approximately true, as in the case of a population of counters under our strict control, valid inferences could be drawn regarding the number of individuals and the death-rate in the population. In the present paper the chief emphasis has been upon testing the truth of the assumptions as they apply in nature. Unless such investigations can be undertaken, the parameters of natural populations cannot be estimated with any confidence from data obtained by means of the capture-recapture method.

The present field data were collected in a manner which avoided as far as possible the grosser errors, such as non-random marking and resampling, or interference with natural processes. No trapping technique, however, can sample a population in a manner that is independent of the behaviour of the individuals, and biased sampling may be unavoidable if there is no homogeneity in the response to the traps. Because this difficulty was encountered, no parameters could be estimated for the *Microtus* population as a whole. However, from several lines of statistical evidence it appeared that once the animals had been trapped they behaved thereafter with sufficient uniformity for valid inferences to be drawn about the marked population. Thus, although the theoretical methods of estimation developed in the previous papers cannot be applied in their entirety to the *Microtus* population, some important features in the latter may still be described. There is also much additional information which we hope to use in a later paper. By contrast, the results for the *Clethrionomys* population provide no evidence by which to reject the necessary assumptions of the marking-recapture method of analysis. Accordingly, the death-rates, dilution-rates and total numbers may be given with some confidence. Even here, however, one reservation must be made.

Let us first of all review the evidence on which the estimates of the total *Microtus* population were rejected and those of *Clethrionomys* were accepted. We knew from the isolated nature of its habitat that the *Microtus* population could not have been recruited by immigration, and from laboratory examinations of other populations in the neighbourhood that no young were being added between October 1948 and April 1949. This absence of birth and immigration during the winter therefore enabled us to recognize that the apparent dilution of a population which was, in fact, decreasing was because the assumptions in the mathematical model were not valid for this population. There were no such objections to the *Clethrionomys* data during either of the non-breeding seasons; but similar evidence cannot be provided for the rest of the year. Parameters for the breeding seasons are therefore estimated for this species on the assumption that the relevant features of behaviour in winter were the same at other times. Whether or not this was the case we have no means of knowing. The estimated population changes, however, seem entirely plausible: a rather slow increase in the numbers of trappable age during the breeding seasons; a fairly large influx at the end, and evidence that these latter animals were at risk of capture throughout their post-weaning existence.

A point of general interest is the following: when a population is subject to birth and immigration throughout the period of sampling, is it possible to prove the validity of the assumptions necessary to estimate its parameters? Let us suppose, for example, that there had been a moderate amount of breeding by *Microtus* in the first winter: would we then have known enough to reject the estimates of population density given in Table 3? The problem of verifying the assumptions might not be solved if, in such a case, the field observations consisted solely of capture-recapture data, and a more elaborate experimental design might be required. It might occasionally be possible to enumerate the entire population of some species at certain times, so that a few of the estimated values could be compared with those observed. Perhaps a more practicable suggestion is to try sampling by methods which depend upon other properties than those which determine the composition of the catch by the methods normally employed. These alternative methods might be too laborious for regular use and would then be used solely as a check. On the other hand, if they proved that the old methods were biased there would be no option but to make a complete change in the field techniques.

Finally, it should be emphasized that the results of this analysis do not imply a universal law for each of the two species. The data for some other *Clethrionomys* population might not prove so satisfactory as those presented here, and for the present, at any rate, it will be necessary to consider each particular case in detail before deciding that these methods of estimation are applicable.

5. SUMMARY

1. Two populations of small rodents, *Microtus agrestis* and *Clethrionomys glareolus*, living on the same area, were sampled by means of a live-trapping technique over a period of very nearly two years. The resulting data are analysed here in some detail in order to see whether any valid estimates of population parameters can be made by the theoretical methods described in two earlier papers. These methods are based on the assumptions that the sampling of the population is entirely at random and that all classes of marked and unmarked animals are caught with equal facility. The main purpose of the present paper is therefore to see whether or not these assumptions were true in the field.

2. The sampling of the *Microtus* population as a whole proved unsatisfactory, since it appeared that during the winter months marked and unmarked animals had not the same chance of being trapped, members of the latter class evidently being less willing to enter the traps. Since this difference in behaviour was observed during the non-breeding season, when no new members should have been entering the population at risk of capture, we could not exclude the possibility that it was also occurring at other seasons of the year. In consequence, it was not possible to estimate the total numbers of *Microtus* on the area by methods which are based on an assumed random sampling of all classes in the population.

3. Once trapped, however, the individuals in the *Microtus* population appeared to behave satisfactorily from the point of view of capture-recapture theory. It was therefore possible to use the data for these marked animals for the estimation of the death-rate, and for calculating the further expectation of life of animals marked for the first time and released in different months. Such figures, however, which can be estimated by the methods illustrated in the text, can only refer, strictly speaking, to that part of the total population which had been marked.

4. The *Clethrionomys* population, on the other hand, behaved differently. There was no evidence of any differential sampling of the various classes of marked and unmarked animals, as in the *Microtus* population. Consequently, we can apply to these data, with some degree of confidence, the appropriate methods of estimating the various population parameters.

The expenses of this investigation were partly paid for by a grant from the Agricultural Research Council.

APPENDIX

The method of obtaining approximate estimates of N_t and P_t from a long chain of samples grouped by Method B, which was described in II, § 5*b*, was modified in the following way. Instead of using only the last two entries, $m_{t-2,t}$ and $m_{t-1,t}$, in each column of the original table, we construct a new table by retaining each value of $m_{t-1,t}$ and forming the sums $\sum_{x=0}^{t-2} m_{xt}$, which we define as n_t $(t = 2, 3, 4, \ldots, T)$. Including the unmarked animals, we have therefore at each sampling subsequent to $t = 1$ the three classes u_t, n_t and $m_{t-1,t}$ $(n_t + m_{t-1,t} = s_t;\ s_t + u_t = C_t)$. Since ψ_t is the total number of marked individuals in the population as a whole at time t,

$$\left.\begin{aligned} \psi_0 &= 0, \\ \psi_{t+1} &= P_t(\psi_t + y_t), \end{aligned}\right\} \tag{1}$$

where, allowing for the d_t accidental losses or removals, $y_t = u_t - d_t = R_t - s_t$. Then we have the following expected number of individuals in the population as a whole, together with the observed numbers in the three classes, at the time the sample was taken at $t+1$:

$$\begin{array}{cc} N_{t+1} - \psi_{t+1} & u_{t+1} \\ \psi_{t+1} - P_t R_t & n_{t+1} \\ P_t R_t & m_{t,t+1} \\ \hline N_{t+1} & C_{t+1} \end{array}$$

It is convenient to introduce a new parameter W_t defined by

$$N_{t+1} = P_t W_t,$$

and from (1) to rewrite this table of expected and observed numbers in the form

$$\begin{array}{cc} P_t[W_t - (\psi_t + y_t)] & u_{t+1} \\ P_t(\psi_t - s_t) & n_{t+1} \\ P_t R_t & m_{t,t+1} \\ \hline P_t W_t & C_{t+1} \end{array}$$

from which we may write down the log likelihood equation, after cancelling P_t,

$$L = u_{t+1} \log [W_t - (\psi_t + y_t)] + n_{t+1} \log (\psi_t - s_t) - C_{t+1} \log W_t. \tag{2}$$

We may now obtain the maximum-likelihood estimates $\hat{W}_t$ and $\hat{\psi}_t$, together with their variances and covariance, in the usual way. Since, however, we are concerned here not so much with ψ_t and W_t, as with other parameters which are related to them, it is simpler to proceed in another fashion. It will be seen that we have two degrees of freedom for estimating these two parameters and that the maximum-likelihood equations are independent. Under these conditions we can equate expected and observed numbers (Bailey, 1951). Thus, in the present example, we have the pair of equations

$$n_{t+1} = C_{t+1}(\psi_t - s_t)/W_t,$$

$$m_{t,t+1} = C_{t+1} R_t / W_t,$$

from which

$$\hat{\psi}_t = \frac{n_{t+1} R_t}{m_{t,t+1}} + s_t \quad (t = 1, 2, 3, \ldots, T-1) \tag{3}$$

and

$$\hat{W}_t = R_t C_{t+1} / m_{t,t+1}.$$

This last equation may be rewritten with the help of (3), using

$$s_t = R_t - y_t \quad \text{and} \quad n_{t+1} + m_{t,t+1} = s_{t+1},$$

in the form

$$\hat{W}_t = (\hat{\psi}_t + y_t)\, C_{t+1}/s_{t+1} \quad (t = 0, 1, 2, \ldots, T-1). \tag{4}$$

Then, from (1),

$$\hat{P}_t = \hat{\psi}_{t+1}/(\hat{\psi}_t + y_t) \quad (t = 0, 1, 2, \ldots, T-2), \tag{5}$$

and since $N_t = P_{t-1} W_{t-1}$, we have, from (4) and (5),

$$\hat{N}_t = \hat{\psi}_t C_t / s_t \quad (t = 1, 2, 3, \ldots, T-1). \tag{6}$$

The variances and covariances of these estimates, appropriate to large-scale sample theory, can be obtained by means of the type of formula given by Fisher (1950, § 55). Thus, if ϕ_1 and ϕ_2 are two functions of the observations a_i, of which the expectations are m_i, with $\Sigma a_i = \Sigma m_i = n$, the total size of the sample,

then
$$V(\phi_1) = \sum_i m_i \left(\frac{\partial \phi_1}{\partial a_i}\right)^2 - n\left(\frac{\partial \phi_1}{\partial n}\right)^2, \tag{7}$$

and a similar equation for $V(\phi_2)$, while the covariance is

$$\operatorname{cov}(\phi_1, \phi_2) = \sum_i m_i \left(\frac{\partial \phi_1}{\partial a_i}\right)\left(\frac{\partial \phi_2}{\partial a_i}\right) - n\left(\frac{\partial \phi_1}{\partial n}\right)\left(\frac{\partial \phi_2}{\partial n}\right).$$

In applying these formulae to the present type of problem, there are, however, two points to be borne in mind. In the first place we may have a parameter which is a function of the observations from more than one sample, in which case we must also sum over the samples. And, secondly, it is necessary to distinguish between observations which enter into the expectations as given values, and the actual observations themselves. Thus, in equation (3) it is true that R_t and s_t are observed values at time t, but they enter into this expression as part of the expectations at $t+1$, which are conditional on the results of previous sampling. Here we regard ψ_t as a function of the observations at $t+1$, and since C_{t+1} does not occur explicitly, there is no term in $V(\psi_t)$ corresponding to $n(\partial\phi_1/\partial n)^2$ in (7). Thus from (7) we have, in terms of observed values,

$$V(\psi_t) = \frac{n_{t+1}R_t^2}{(m_{t,t+1})^2} + \frac{n_{t+1}^2 R_t^2}{(m_{t,t+1})^3} = \frac{R_t^2 n_{t+1} s_{t+1}}{(m_{t,t+1})^3}.$$

Formally, we now should substitute expected for observed values, and write

$$V(\psi_t) = \frac{W_t(\psi_t + y_t)(\psi_t - s_t)}{R_t C_{t+1}},$$

a result which may be verified by the more lengthy process of forming the maximum-likelihood equations from (2) for ψ_t and W_t, and then inverting the corresponding information matrix. In actual practice, of course, we would calculate the numerical value of $V(\psi_t)$ by inserting the estimated $\hat{W}_t$ and $\hat{\psi}_t$ in this expression, and since W_t is a parameter which was introduced merely for convenience (it is not necessary here to calculate its actual value), it is more practical to write

$$V(\psi_t) = (\psi_t + y_t)^2 \frac{n_{t+1}}{s_{t+1} m_{t,t+1}}.$$

In a similar way we have
$$V(P_t) = P_t^2\left[\frac{V(\psi_t)}{(\psi_t + y_t)^2} + \frac{V(\psi_{t+1})}{\psi_{t+1}^2}\right],$$

and
$$V(N_t) = N_t^2\left[\frac{C_t - s_t}{s_t C_t} + \frac{V(\psi_t)}{\psi_t^2}\right].$$

There is one further parameter we wish to determine which arises in the analysis of the subpopulation of marked animals, namely, Z_t, the number of new additions to this population which were made at time t. From the theoretical point of view we therefore consider a population consisting of a variable number N_t of individuals at the time the sample of C_t is withdrawn, and we make the rule that any dilution of this population with new members shall only take place at the actual time a sample was returned. Thus, if an unknown number of Z_t new members enter at time t, when the R_t survivors out of the original sample of C_t are returned, the total number of individuals in the population at this moment is $N_t + Z_t - d_t$. Then, since by definition no further new members enter during the interval t to $t+1$, the total number of individuals when the sample of C_{t+1} is taken at $t+1$ is

$$N_{t+1} = P_t(N_t + Z_t - d_t).$$

Now, clearly $(N_t + Z_t - d_t)$ is the same parameter as we have defined earlier as W_t, and hence from (4) and (6)

$$\hat{Z}_t = \frac{(\hat{\psi}_t + y_t)\,C_{t+1}}{s_{t+1}} - \frac{\hat{\psi}_t C_t}{s_t} + d_t, \tag{8}$$

or alternatively
$$\hat{Z}_t = R_t\left[\frac{s_t C_{t+1} - n_{t+1} C_t}{s_t m_{t,t+1}} - 1\right] \quad (t = 1, 2, 3, \ldots, T-1).$$

The latter form is more convenient in practice if we merely require the estimates for a set of observed frequencies, but if in addition we also require the variances of these estimates, then the former is to be preferred. For, if we write $p_t = s_t/C_t$, $q_t = 1-p_t$, we have, proceeding as before,

$$V(\hat{Z}_t) = \left(\frac{\hat{\psi}_t + y_t}{p_{t+1}}\right)^2 \left\{\frac{1}{s_{t+1}}\left[\left(\frac{p_t - p_{t+1}}{p_t}\right)^2 \frac{n_{t+1}}{m_{t,t+1}} + q_{t+1}\right]\right\} + \left(\frac{\hat{\psi}_t}{p_t}\right)^2 \frac{q_t}{s_t}.$$

All these results can be verified by considering the expected numbers in the population and the observed numbers in the samples taken at t and $t+1$. Thus we have

$$\begin{array}{cccc}
\multicolumn{2}{c}{t} & \multicolumn{2}{c}{t+1} \\
N_t - \psi_t & u_t & P_t[W_t - (\psi_t + y_t)] & u_{t+1} \\
\psi_t - P_{t-1}R_{t-1} & n_t & P_t(\psi_t - s_t) & n_{t+1} \\
P_{t-1}R_{t-1} & m_{t-1,t} & P_t R_t & m_{t,t+1} \\
\hline
N_t & C_t & P_t W_t & C_{t+1}
\end{array}$$

from which we can easily write down the maximum-likelihood equations for the simultaneous estimation of the four parameters which can be determined from these samples, namely, P_{t-1}, ψ_t, N_t and W_t, and obtain the corresponding 4×4 information matrix. By inverting the latter, the variances and co-variances are obtained, and finally $\hat{Z}_t$ and $V(\hat{Z}_t)$ from

$$\hat{Z}_t = \hat{W}_t - \hat{N}_t + d_t,$$

$$V(\hat{Z}_t) = V(\hat{W}_t) + V(\hat{N}_t) - 2\operatorname{cov}(\hat{W}_t, \hat{N}_t).$$

This procedure is, however, very much more lengthy and tedious than the one adopted here.

There remains the question whether this maximum-likelihood estimate of Z_t is biased. It will be noted from (8) that $\hat{Z}_t$ is essentially the difference between two estimates, each of which is of a type known to be positively biased (II, § 1). If we could assume that the absolute amount of bias in each case was of much the same order, then we might regard $\hat{Z}_t$ are being approximately unbiased. But it is doubtful as to how far this assumption would be true in general, and it is more likely that when $Z_t > 0$ the bias of the positive term would tend to be greater than that of the negative. Under these circumstances $\hat{Z}_t$ might have some degree of positive bias, and in place of (8) we might therefore prefer to use the adjusted estimate,

$$\breve{Z}_t = \frac{(\psi_t + y_t)(C_{t+1}+1)}{s_{t+1}+1} - \frac{\psi_t(C_t+1)}{s_t+1} + d_t.$$

Since the comparison of $\hat{Z}_t$ with the known true values of this parameter played such an important part in the argument developed in the text, it was therefore of interest to see whether the use of this adjusted estimate would have made any difference to our conclusions. The following were the comparative results obtained in the case of the data for the marked *Microtus* population:

	True values	$\hat{Z}_t$	$\breve{Z}_t$	$\Delta(\hat{Z}_t)$	$\Delta(\breve{Z}_t)$
Sept.	65	48·2	44·8	−16·8	−20·2
Oct.	41	51·4	46·5	+10·4	+ 5·5
Nov.	45	38·9	37·5	− 6·1	− 7·5
Mar.	76	66·8	65·3	− 9·2	−10·7
Apr.	65	111·8	107·3	+46·8	+42·3
May	46	19·6	22·3	−26·4	−23·7
June	48	35·5	35·2	−12·5	−12·8
July	54	69·5	67·4	+15·5	+13·4
Sept.	65	78·8	78·4	+13·8	+13·4
Nov.	108	88·9	81·3	−19·1	−26·7
Total	613	609·4	586·0		

It will be seen that, with one exception, $\check{Z}_t < \hat{Z}_t$, and that we do not obtain quite such a close correspondence between the estimated and known total number of *Microtus* released over the whole period. Taking the individual $\check{Z}_t$, however, it seems that on the average they are, if anything, in slightly better agreement with the true values than the $\hat{Z}_t$. Thus, the sum of the squares of the deviations $\Sigma\Delta^2(\check{Z}_t) = 4196$, whereas $\Sigma\Delta^2(\hat{Z}_t) = 4351$. These differences are, however, slight. We concluded, therefore, that if the maximum-likelihood estimate $\hat{Z}_t$ was biased, the degree of this bias was probably small compared with the sampling errors involved, at any rate in this series of data.

REFERENCES

BAILEY, N. T. J. (1951). Testing the solubility of maximum likelihood equations in the routine application of scoring methods. *Biometrics*, **7**, 268–74.

CHITTY, D. (1952). Mortality among voles (*Microtus agrestis*) at Lake Vyrnwy, Montgomeryshire in 1936–9. *Phil. Trans.* B, **236**, 505–52.

CHITTY, D. & KEMPSON, D. A. (1949). Prebaiting small mammals and a new design of live trap. *Ecology*, **30**, 536–42.

FISHER, R. A. (1950). *Statistical Methods for Research Workers*, 11th ed. Edinburgh: Oliver and Boyd.

LESLIE, P. H. (1952). The estimation of population parameters from data obtained by means of the capture-recapture method. II. The estimation of total numbers. *Biometrika*, **39**, 363–88.

LESLIE, P. H. & CHITTY, D. (1951). The estimation of population parameters from data obtained by means of the capture-recapture method. I. The maximum-likelihood equations for estimating the death-rate. *Biometrika*, **38**, 269–92.

ON THE UTILIZATION OF MARKED SPECIMENS IN ESTIMATING POPULATIONS OF FLYING INSECTS

BY C. C. CRAIG
University of Michigan

1. INTRODUCTION

Professor William Hovanitz called my attention to the following problem: An observer catches butterflies, marks them, and immediately releases them. It is assumed that a butterfly, no matter how many times it has been caught before, has the same susceptibility to capture as any other butterfly in the population which is supposed stable while the captures are being made. Records are kept of f_x, the frequency of cases in which the same butterfly is caught x times, $x = 1, 2, \ldots$, until a total of s captures of r different butterflies have been made. ($\Sigma f_x = r$; $\Sigma x f_x = s$.) The number, f_0, of butterflies which escape is not observed; the problem is to estimate from the values of f_x the total population n of butterflies on the area assumed well defined.

The estimation of biological populations by means of capture-recapture data is by no means a new problem, though papers dealing with it from the mathematical-statistical point of view are largely quite recent. (In particular, see the papers by Leslie & Chitty (1951), Bailey (1951), Moran (1951, 1952), and the bibliographies quoted by them.) However, the experimental conditions and the mathematical models for the present study appear to differ in essential ways from those previously considered. The important point of departure is that each butterfly on being netted is immediately marked (with a spot of nail polish) and released. The butterflies (*Colias eurytheme*) were caught in one of two isolated alfalfa fields, which they inhabit, in southern California. Each catch was made during the same day at times when the butterflies were freely flying. Thus, it seemed reasonable to assume that the population was stable during a catch. The experimenter, Prof. Hovanitz, endeavoured to give each butterfly an equal chance of capture, walking in straight lines across the field and deviating in direction before reaching a boundary only when he noticed that a butterfly just caught tended to fly down his path. One check of the suitability of a mathematical model is to test the agreement of the experimental results with respect to the number of butterflies caught once, twice, etc., with those predicted from the model. I will return to this point at the end of the paper.

Two mathematical models seem appropriate to serve as a basis for discussion of this estimation problem. It is of some interest to see that both lead to approximately the same estimates with little difference in their precision for large samples. It may be of more interest that for both models in which the population size is regarded as a parameter, though maximum likelihood estimates exist and agree substantially with moment estimates in all sixteen of the actual field experiments for which I have data, nevertheless with increasing sample size meaningful solutions of the likelihood equation do not exist.

2. THE APPLICATION OF A TRUNCATED POISSON DISTRIBUTION

If we suppose that the total period during which captures are made is composed of a large number of short intervals sufficient for the capture of a butterfly and that each butterfly is equally subject to being netted in each such interval, it appears that one may consider that

$$E(f_x) = n\,e^{-\lambda}\frac{\lambda^x}{x!} \quad (x = 0, 1, 2, \ldots).$$

If so, the likelihood of the observed sample is

$$L = \frac{e^{-n\lambda}\lambda^s n!}{(1!)^{f_1}(2!)^{f_2}\ldots(n-r)!\,f_1!\,f_2!\ldots}, \tag{1}$$

in which n and λ are parameters.

Before proceeding to maximum-likelihood estimates, we may note that two moment estimates are immediately available.

Method 1. We have

$$E(r) = n(1-e^{-\lambda}),$$

$$E(s) = n\lambda.$$

Equating the observed r and s to their expected values, we estimate λ from

$$\frac{r}{s} = \frac{1-e^{-\lambda}}{\lambda}. \tag{2}$$

Using this value of λ, we estimate n from

$$\tilde{n} = s/\lambda. \tag{3}$$

In practice we can estimate n directly by eliminating λ between (2) and (3), getting

$$\log n - \log(n-r) = s/n,$$

which is readily solved for the integer n which most nearly satisfies it by the use of a good table of natural logarithms.

In correspondence with Prof. Hovanitz, Prof. Sewall Wright gave the above estimate and also suggested the use of the method of maximum likelihood.

Method 2. It is still simpler to use the first and second power sums

$$s\,(=s_1) \quad \text{and} \quad s_2 = \Sigma x^2 f_x.$$

For a Poisson distribution,

$$E(s/n) = E(s_2/n) - [E(s/n)]^2,$$

or

$$[E(s)]^2 = n[E(s_2) - E(s)].$$

Thus using s and s_2 for their expected values we have the estimate,

$$\tilde{n} = s^2/(s_2 - s).$$

Though this is obviously subject to greater sampling error than the first estimate and the others to follow, it serves as a quickly obtained rough value which also can be used as a first trial value in solving the other estimation equations.

Method 3. Let us consider the truncated Poisson distribution for which we have observed frequencies as a complete distribution function, i.e. we write

$$f_x = \frac{e^{-\lambda}}{1-e^{-\lambda}}\frac{\lambda^x}{x!} \quad (x = 1, 2, \ldots).$$

Now
$$E\left(\frac{s}{n}\right) = E(\bar{x}) = \frac{\lambda e^{\lambda}}{e^{\lambda}-1}, \tag{4}$$

and equating this to s/r, we estimate λ from
$$r/s = (1-e^{-\lambda})/\lambda,$$
just as in method 1.

But now the likelihood of the sample is
$$L = \frac{\lambda^{s} r!}{(e^{\lambda}-1)^{r}(1!)^{f_1}(2!)^{f_2}\ldots f_1!f_2!\ldots},$$
and
$$\log L = s\log\lambda - r\log(e^{\lambda}-1) + \text{terms free of } \lambda.$$
Then
$$\frac{d\log L}{d\lambda} = \frac{s}{\lambda} - \frac{r}{1-e^{-\lambda}} = 0$$
is equation (4) again. Thus in this case the estimates obtained by the moments principle and by maximum likelihood agree. We next estimate n from the relation
$$E(r) = n(1-e^{-\lambda}),$$
as in method 1. That is, methods 1 and 3 are mechanically the same, but under the argument for method 3 the estimate for λ is that given by maximum likelihood.

Method 4. We return to the likelihood function (1) in which both n and λ are regarded as parameters. Now
$$\log L = -n\lambda + s\log\lambda + \log[n^{(r)}] + \text{terms free of } \lambda \text{ and } n$$
and
$$\frac{\partial\log L}{\partial\lambda} = -n + \frac{s}{\lambda}, \quad \frac{\partial\log L}{\partial n} = -\lambda + \sum_{k=0}^{r-1}\frac{1}{n-k}.$$
Equating these to zero and eliminating λ, the estimation equation for n is
$$\frac{s}{n} = \sum_{k=0}^{r-1}\frac{1}{n-k}.$$
Now the only solutions of this equation of interest are those for which $n \geqslant r$. It is not difficult to show that such solutions exist if
$$s/r \leqslant 1 + \tfrac{1}{2} + \ldots + 1/r.$$
But for any fixed population, $r \leqslant n$, while s, the total number of butterflies caught, may be increased without limit (ignoring the wear and tear on the butterflies). Thus for increasing sample size the likelihood equation fails to have a meaningful solution, and the usual theorems on the asymptotic behaviour of likelihood estimates are not available. Of course as $s \to \infty$, r has n for its stochastic limit. For all the sixteen field experiments for which I have data, solutions $n > r$ do exist and can be found using tables of sums of reciprocals for n's up to 450 by using the estimate given by method 2. For larger n's a good approximation to $\sum_{k=0}^{r-1} 1/(n-k)$ is given by
$$\log n - \log(n-r+1) + \tfrac{1}{2}[1/(n-r+1) + 1/n],$$
and the estimation relation on neglecting terms of $O(n-r+1)^{-2}$ is
$$\log n - \log(n-r+1) + \tfrac{1}{2}[1/(n-r+1) + 1/n] = s/n.$$
As shown in the table at the end, these estimates agree quite well with those given by the other methods.

3. THE APPLICATION OF STEVENS'S DISTRIBUTION FUNCTION FOR GROUPS

A second mathematical model may be arrived at by deriving an expression for the expected number of times a butterfly will be caught if a total of s captures is made, if it is assumed that each butterfly, no matter how many times it has been netted before, is equally liable to capture and by considering after each capture the probability that the next will be of a butterfly that has been caught before. But this was seen to be equivalent to a problem solved by W. L. Stevens (1937) as follows: Let each of a set of objects on being drawn be equally likely to be assigned to any one of n classes. After s objects have been drawn and so assigned to classes at random, what is the probability that exactly r classes will be occupied? Stevens's solution (which I also obtained before consulting his paper) is

$$u_{rs} = \frac{n^{(r)}}{n^s}\frac{\Delta^r 0^s}{r!} = \frac{n^{(r)}}{n^s}\sigma_s^r \tag{5}$$

in which σ_s^r is a Stirling's number of the second kind (see Jordan, 1947, pp. 171–2). (Actually Stevens's problem was not original with him either; see Jordan, 1947, pp. 177–9.)

With this distribution law n is the only parameter to be estimated. Again this can be found from a sample with a given r and s either by moments or by maximum likelihood.

Method 5. Stevens found that

$$E(r) = n[1-(1-1/n)^s],$$

and thus from a single sample we have the estimation relation

$$1-r/n = (1-1/n)^s.$$

This can be solved by trial and error by means of a good table of logarithms fairly conveniently by rewriting it

$$\log(n-r) = s\log(n-1)-(s-1)\log n. \tag{6}$$

Method 6. The likelihood of the sample is given by (5). We have

$$\log L = \log u_{rs} = \log n^{(r)} - s\log n + \text{terms free of } n,$$

and

$$\frac{d\log L}{dn} = \sum_{k=0}^{r-1}\frac{1}{n-k}-\frac{s}{n}. \tag{7}$$

This equated to zero gives the same estimation relation as in method 4 though arrived at from a different model. That this is to be expected can be verified by observing that if in (1) λ is replaced by s/n, then to maximize (1) is to maximize $n^{(r)}/n^s$.

Moreover, it is readily seen that numerically methods 5 and 6 are almost equivalent. In practice we take in method 6 the value of n for which the right member of (7) is most nearly zero. Suppose that instead we seek the largest value of n for which

$$\frac{(n-1)^{(r)}}{(n-1)^s} \leqslant \frac{n^{(r)}}{n^s}.$$

But this inequality reduces to

$$1-\frac{r}{n} \leqslant \left(1-\frac{1}{n}\right)^s.$$

However, it is obvious that in this case, too, there are the same difficulties in regard to the large sample behaviour of this estimator as in method 4.

Let us now derive the large sample variances of the other estimates which are three in number, since methods 1 and 3 are equivalent.

For method 1 we proceed as follows: With λ replaced by its estimate s/n, the estimation relation is

$$r = n(1-e^{-s/n}),$$

and from this

$$\frac{\delta n}{n} = \frac{\delta r - e^{-\lambda}\,\delta s}{n(1-e^{-\lambda}-\lambda\,e^{-\lambda})}. \tag{8}$$

Of course $\sigma_{\tilde{n}}$, if $\tilde{n}$ is the estimate, will be proportional to n, and we will find $\sigma^2_{\tilde{n}/n}$ which is given to $O(1/n)$ by finding the expected value of the square of the right member of (8). We have then to $O(1/n)$

$$\sigma^2_{\tilde{n}/n} = \frac{e^{-\lambda}(1-e^{-\lambda})-2e^{-\lambda}\,\lambda\,e^{-\lambda}+\lambda\,e^{-2\lambda}}{n(1-e^{-\lambda}-\lambda\,e^{-\lambda})^2} = \frac{1}{n(e^{\lambda}-1-\lambda)}. \tag{9}$$

For method 2 proceeding in the same way, starting with $n = s^2/(s_2-s)$, we have

$$\frac{\delta n}{n} = \frac{2(s_2-s)\,s\delta s - s^2(\delta s_2-\delta s)}{n(s_2-s)^2}.$$

Taking the expected value of the square of the right member and the values of moments about zero of a variable obeying the Poisson distribution law, we get the simple result to $O(1/n)$ in this case,

$$\sigma^2_{\tilde{n}/n} = \frac{2}{n\lambda^2}. \tag{10}$$

As is to be expected, this variance is larger than for method 1.

For method 5 the result is a bit more cumbersome. We first get

$$\frac{\delta n}{n} = \frac{(n-1)\,\delta r}{r(s+n-1)-ns},$$

in which on the right r satisfies the relation

$$r = n\left[1-\left(1-\frac{1}{n}\right)^s\right].$$

Next we must find σ^2_r. We have

$$E(r) = n\left[1-\left(1-\frac{1}{n}\right)^s\right],$$

and it is easy to find $\quad E[r(r-1)] = \dfrac{n-1}{n^{s-1}}[n^s-2(n-1)^s+(n-2)^s].$

Using these $\quad \sigma^2_r = n\left(1-\dfrac{1}{n}\right)^s + n(n-1)\left(1-\dfrac{2}{n}\right)^s - n^2\left(1-\dfrac{1}{n}\right)^{2s}.$

This quantity can be so small for even moderately large n's even though the last two terms are large that it appears necessary to calculate it just as it stands. We have to order $O(1/n)$

$$\sigma_{\tilde{n}/n} = \frac{(n-1)\,\sigma_r}{r(s+n-1)-ns}. \tag{11}$$

I will illustrate the use of these various formulae for estimates of n and their variances with a pair of numerical examples from data of Prof. Hovanitz.

Example 1

x	f_x
0	...
1	66
2	3
	$69 = r$

$s = 72, \quad s_2 = 78$

By method 1 (or 3) on solving $\log n - \log (n-69) = 72/n$ we get $n = 840$. Our estimate of λ is then $72/840$, from which

$$\sigma^2_{\tilde{n}/n} = \frac{1}{840(e^{\lambda}-1-\lambda)} = 0{\cdot}315$$

and

$$\sigma_{\tilde{n}/n} = 0{\cdot}561.$$

By method 2,

$$\tilde{n} = \frac{72^2}{78-72} = 864,$$

$$\sigma_{\tilde{n}/n} = \left(\frac{2}{n\lambda^2}\right)^{\frac{1}{2}} = \left(\frac{2n}{s^2}\right)^{\frac{1}{2}} = 0{\cdot}577.$$

By method 4, on solving

$$\log n - \log (n-68) + \frac{1}{2}\left(\frac{1}{n-68} + \frac{1}{n}\right) = \frac{72}{n},$$

we get $n = 828$. By method 5 (or 6) we also get $n = 828$. For method 5 we have

$$\sigma_{\tilde{n}/n} = 0{\cdot}55.$$

All of these values of the standard deviation of the proportional error of the estimate of n are high. This illustrates the fact that for an n of this magnitude a catch of only about 8 % of the population was not an adequate sample for the estimation of the total.

Example 2. Let us see how matters turn out in another example in which the estimated population is about the same but the total catch is some six times greater.

Data

x	f
0	...
1	258
2	72
3	3
	$341 = r$

$s = 435, \quad s_2 = 901$

Estimates and standard errors

Method	$\tilde{n}$	$\sigma_{\tilde{n}/n}$
1 and 3	856	0·0871
2	901	0·0976
4	853	...
5	853	0·0867
6	853	...

The following table is a summary of the results for sixteen experimental catches.

Finally, I return to the question mentioned earlier, of testing the suitability of the models used to the experimental conditions. For the first model using the estimates of n and λ obtained, it is easy to calculate the predicted frequencies for $x = 0, 1, 2, \ldots$. For the two examples given in detail using the estimates found by methods 1 or 3, the frequencies calculated for $x = 0, 1, 2, \ldots$ are: *Example* 1: 771·0, 66·1, 2·8, 0·1; and *Example* 2: 515·1, 261·7, 66·5, 11·2, 1·4, 0·1. A χ^2 test is superfluous for the first; for the second the significance level is above 0·5. I have similarly tested the remaining fourteen examples for which I have data. The significance level for χ^2 was above 0·5 in every case but one, in which it was approximately 0·3. The second model does not lend itself to a similar check. For that I only have to offer the fact that the estimates obtained by its use agree quite well with those found from the first model.

Observed values			Estimates of n by method			Values of $\sigma_{\tilde{n}/n}$ for method		
r	s	s_2	1 and 3	2	4, 5 or 6	1 and 3	2	5
69	72	78	840	864	828	0·56	0·58	0·55
93	108	140	352	364	348	·23	·25	·23
159	187	249	560	564	557	·17	·18	·17
144	161	197	708	720	703	·22	·24	·22
63	76	108	196	180	193	0·24	0·25	0·24
56	66	92	195	168	192	·28	·28	·28
48	74	188	79	86	78	·14	·18	·14
341	435	645	856	901	853	·087	·098	·087
276	330	450	895	908	892	0·11	0·13	0·12
222	249	303	1063	1148	1059	·18	·19	·18
154	225	385	277	316	275	·090	·112	·090
148	180	256	444	426	442	·15	·16	·15
63	91	155	116	129	114	0·15	0·18	0·14
60	98	216	91	81	90	·11	·13	·11
71	118	258	105	99	104	·10	·12	·098
46	89	211	59	65	58	·092	·13	·088

REFERENCES

BAILEY, N. T. J. (1951). On estimating the size of mobile populations from recapture data. *Biometrika*, **38**, 292–306.

JORDAN, CHARLES (1947). *Calculus of Finite Differences*, 2nd ed. New York.

LESLIE, P. H. & CHITTY, DENNIS (1951). The estimation of population parameters from data obtained by means of the capture-recapture method. I. The maximum likelihood equations for estimating the death rate. *Biometrika*, **38**, 269–92.

MORAN, P. A. P. (1951). A mathematical theory of animal trapping. *Biometrika*, **38**, 307–11.

MORAN, P. A. P. (1952). The estimation of death rates from capture-mark-recapture sampling. *Biometrika*, **39**, 181–8.

STEVENS, W. L. (1937). The significance of grouping. *Ann. Eugen., Lond.*, **8**, 57–69.

THE TOTAL SIZE OF A GENERAL STOCHASTIC EPIDEMIC

By NORMAN T. J. BAILEY
Nuffield Lodge, Regent's Park, London

1. Introduction

The early work on the mathematical theory of epidemics (e.g. Ross,* 1916 and later; Brownlee, 1918; Kermack & McKendrick, 1927 and later; Soper, 1929) was invariably of a 'deterministic' nature, and assumed that for given numbers of susceptible and infectious individuals, and given attack and removal rates, a certain definite number of fresh cases would occur in any specified time. However, it is widely realized that an appreciable element of chance enters into the conditions under which new infections or removals take place. The probability approach was fundamental to Greenwood's (1931, 1946) use of chain binomials in discussing the distribution of multiple cases of disease in households. Recent discussions of these problems (e.g. Bartlett, 1946, 1949; Bailey, 1950) have therefore turned to 'stochastic' models. In these we have, for any given instant of time, probability distributions for the total numbers of susceptible and infected individuals replacing the single point-values of the deterministic treatments. Stochastic models have a special importance in this context due to the fact that for epidemic processes stochastic means are not the same as the corresponding deterministic values. Although for large homogeneously mixing groups deterministic methods might be fairly adequate, it seems likely that in practice epidemics actually occur in several relatively small groups of friends and acquaintances, the epidemiological returns for an administrative unit being compounded of many such comparatively distinct processes. Moreover, when we are considering the distribution of cases of a disease in a household the size of the group is always so small as to demand a stochastic treatment.

In my 1950 paper I discussed a simple stochastic epidemic where none of the infected individuals was removed from circulation by death, recovery or isolation. This might well apply to some of the milder infections of the upper respiratory tract, and can also be used approximately to represent epidemics for which the time taken for removal from circulation is long compared with the time usually required for the epidemic to be completed. The present paper considers the more general problem of allowing for both infection and removal. The analytical difficulties present in the treatment of the simple epidemic appear here in a more acute form, though it has proved possible to compute the frequency distribution of the total size of the epidemic for moderate group size given the ratio of removal to infection rate. The results obtained may be compared with those described by Kermack & McKendrick (1927) in the deterministic case. No obvious analogue to the threshold theorem can be discerned in the stochastic models. An important application of these results is to the problem of the distribution of multiple cases of disease in a household, and a method is given for obtaining maximum-likelihood estimates of the ratio of the removal to infection rate.

---

* Although Ross started with the idea of probability his mathematical theory is essentially deterministic.

2. DETERMINISTIC TREATMENT

We must first glance briefly at the results obtained in the deterministic case. The following treatment, with constant infection and removal rates, is substantially that given by Kermack & McKendrick (1927), though with some slight alterations to their notation.

Consider a homogeneously mixing community of n individuals, of whom at time t there are x susceptibles, y infectious cases in circulation and z individuals who are isolated, dead, or recovered and immune. Thus we have

$$x+y+z = n.$$

Now suppose that there is a constant infection rate β and a constant removal rate γ, so that the number of new infections in time dt is $\beta xy\,dt$ and the number of removals from circulation is $\gamma y\,dt$. We can choose our time scale so that t is replaced by βt. Then it is easy to see that the course of the epidemic is represented by the differential equations

$$\left.\begin{aligned} \frac{dx}{dt} &= -xy, \\ \frac{dy}{dt} &= xy-\rho y, \\ \frac{dz}{dt} &= \rho y, \end{aligned}\right\} \tag{1}$$

where $\rho = \gamma/\beta$, the ratio of the removal to infection rate. Initially, when $t = 0$, we can assume that x is approximately equal to n. It is then clear from (1) that unless $\rho < n$ no epidemic can start to build up as this requires $[dy/dt]_{t=0} > 0$. Kermack & McKendrick obtained an approximate solution to (1) for epidemics of small magnitude and showed that, if $\rho = n-\nu$, where ν is small compared with n, an epidemic of total size 2ν will occur. This constitutes what may be called Kermack & McKendrick's Threshold Theorem, and can be interpreted by saying that if the initial density of susceptibles is $n = \rho+\nu$ then the introduction of a few infected persons will give rise to an epidemic, after which the density of susceptibles is reduced to $\rho-\nu$, a value as far below the threshold ρ as originally it was above it. Somewhat similar results can be obtained for the more general case of variable infection and removal rates, and an extension can be made to the situation where an intermediate host is involved (Kermack & McKendrick, 1927).

3. STOCHASTIC TREATMENT

Let us now consider the stochastic analogue of the deterministic treatment discussed in the previous section. We shall use the same definitions of x, y and z, and shall replace t by βt as before. Then on the assumption of homogeneous mixing of the susceptibles and infectious individuals in circulation the probability of one new infection taking place in time dt is $xy\,dt$, while the probability of one infected person being removed from circulation in time dt is $\rho y\,dt$. Let $p_{rs}(t)$ be the probability that at time t there are r susceptibles still uninfected and s infectious individuals in circulation. Let us assume that the epidemic is started by the introduction of a infectious cases into a population of n susceptibles. It is now easy to show by the usual methods that the whole process can be characterized by the partial differential equation for the probability generating function Π:

$$\frac{\partial \Pi}{\partial t} = (v^2 - uv)\frac{\partial^2 \Pi}{\partial u\, \partial v} + \rho(1-v)\frac{\partial \Pi}{\partial v}, \tag{2}$$

where

$$\Pi = \sum_{r,s} u^r v^s p_{rs}, \tag{3}$$

with limits

$$0 \leqslant r+s \leqslant n+a, \quad 0 \leqslant r \leqslant n, \quad 0 \leqslant s \leqslant n+a. \tag{4}$$

Equation (2) is substantially that given by Bartlett (1949, equation (49)), putting his immigration rate equal to zero.

Let us now use the Laplace transform and its inverse with respect to time given by

$$\left.\begin{aligned} \phi^*(\lambda) &= \int_0^\infty \mathrm{e}^{-\lambda t}\phi(t)\, dt, \quad \mathbf{R}(\lambda) > 0, \\ \phi(t) &= \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{\lambda t}\phi^*(\lambda)\, d\lambda, \end{aligned}\right\} \tag{5}$$

where $\int_{c-i\infty}^{c+i\infty} \equiv \lim_{w\to\infty}\int_{c-iw}^{c+iw}$, and c is positive and greater than the abscissae of all the residues. Taking transforms of (2) and (3), and using the boundary condition

$$p_{na}(0) = 1, \tag{6}$$

we obtain

$$(v^2 - uv)\frac{\partial^2 \Pi^*}{\partial u\, \partial v} + \rho(1-v)\frac{\partial \Pi^*}{\partial v} - \lambda \Pi^* + u^n v^a = 0, \tag{7}$$

and

$$\Pi^* = \sum_{r,s} u^r v^s p^*_{rs} = \sum_{r,s} u^r v^s q_{rs}, \tag{8}$$

where

$$q_{rs} = p^*_{rs} = \int_0^\infty e^{-\lambda t} p_{rs}(t)\, dt. \tag{9}$$

Substituting (8) in (7), and equating coefficients of $u^r v^s$, yields the recurrence relations

$$\left.\begin{aligned} &(r+1)(s-1)\,q_{r+1,s-1} - \{s(r+\rho)+\lambda\}\,q_{rs} + \rho(s+1)\,q_{r,s+1} = 0 \\ \text{and} \quad &-\{a(n+\rho)+\lambda\}\,q_{na} + 1 = 0, \\ \text{with} \quad &0 \leqslant r+s \leqslant n+a, \quad 0 \leqslant r \leqslant n, \quad 0 \leqslant s \leqslant n+a. \end{aligned}\right\} \tag{10}$$

Any q_{rs} whose suffix falls outside the prescribed ranges is taken to be identically zero. It is evident from the form of the equations that, starting with q_{na}, all the quantities q_{rs} could be calculated successively. Using the inverse of the Laplace transformation, we could then arrive at the required p_{rs}, exhibiting them as sums of exponential terms like $e^{-i(j+\rho)t}$. There seems to be considerable difficulty in handling such expressions in a compact and convenient way to give, for example, epidemic completion times or the stochastic epidemic curve showing the rate of change with respect to time of the average total number of removals at any instant. However, some progress is possible if we concentrate attention on the total size of the epidemic, i.e. the value of $n+a-x$ for $t=\infty$. As $t\to\infty$ all terms in the expansion of $p_{rs}(t)$ involving negative exponentials like $e^{-i(j+\rho)t}$ vanish unless $i=0$. The non-vanishing term is the coefficient of λ^{-1} in the partial fraction expansion of q_{rs} in terms of $\{i(j+\rho)+\lambda\}^{-1}$.

Now the epidemic ceases to spread to fresh susceptibles as soon as $s = 0$. Thus the probability of an epidemic of total size w (not counting the initial a infectious persons) is

$$
\begin{aligned}
P_w &= \lim_{t\to\infty} p_{n-w,0}(t) \quad (0 \leqslant w \leqslant n), \\
&= \lim_{\lambda\to 0} \lambda q_{n-w,0} \\
&= \lim_{\lambda\to 0} \rho q_{n-w,1}, \quad \text{putting } r = n-w \text{ and } s = 0 \text{ in (10)}, \\
&= \rho f_{n-w,1}, \qquad (11)
\end{aligned}
$$

where

$$
\left.
\begin{aligned}
&f_{rs} = \lim_{\lambda\to 0} q_{rs}, \\
&1 \leqslant r+s \leqslant n+a, \quad 0 \leqslant r \leqslant n, \quad 1 \leqslant s \leqslant n+a.
\end{aligned}
\right\} \qquad (12)
$$

for

The quantities f_{rs} evidently satisfy the following recurrence relations obtained from (10) by writing f_{rs} for q_{rs} and putting $\lambda = 0$,

$$
\left.
\begin{aligned}
(r+1)(s-1)f_{r+1,s-1} - s(r+\rho)f_{rs} + \rho(s+1)f_{r,s+1} &= 0 \\
-a(n+\rho)f_{na} + 1 &= 0,
\end{aligned}
\right\} \qquad (13)
$$

and

with same limits as in (12).

Some further simplification results from writing

$$
f_{rs} = \frac{n!\,(r+\rho-1)!\,\rho^{n+a-r-s}}{sr!\,(n+\rho)!}\, g_{rs}. \qquad (14)
$$

Substituting in (13) gives

$$
\left.
\begin{aligned}
g_{r+1,s-1} - g_{rs} + (r+\rho)^{-1} g_{r,s+1} &= 0 \\
g_{na} &= 1.
\end{aligned}
\right\} \qquad (15)
$$

and

I am indebted to Dr F. G. Foster for suggesting to me the alternative approach of considering the succession of population states represented by the points (r, s). Thus the progress of the epidemic can be regarded as a random walk from the point (n, a) to the points $(n-w, 0)$ $w = 0, 1, \ldots, n$, with an absorbing barrier at $r = 0$, and where the possible transitions from (r, s) are

$(r,s) \to (r-1, s+1)$, occurring with probability $r/(r+\rho)$,

and

$(r,s) \to (r, s-1)$, occurring with probability $\rho/(r+\rho)$.

Foster's general formula for P_w can now be written down almost immediately simply by considering the sum of the probabilities of all possible paths from (n, a) to $(n-w, 0)$. Thus we have

$$
P_w = \frac{\rho^{a+w}}{\rho+n-w} \frac{\binom{n}{w}}{\binom{n+\rho}{w}} \sum_{\alpha} (\rho+n)^{-\alpha_0}(\rho+n-1)^{-\alpha_1}\ldots(\rho+n-w)^{-\alpha_w}, \qquad (16)
$$

where the summation is over all compositions of $a+w-1$ into $w+1$ parts such that $0 \leqslant \alpha_i \leqslant a+i-1$ for $0 \leqslant i \leqslant w-1$ and $1 \leqslant \alpha_w \leqslant a+w-1$. However, for the purposes of computation there appears to be some advantage, especially if n is at all large, in calculating the quantities P_w from (11), (14) and (15), instead of from (16). The reason for this is that not only is the form of (16) not very suitable for computation, but also the partitional nature of the summation may leave some doubt as to whether all relevant terms have been included in any specific instance. Using (11), (14) and (15), therefore, the P_w have been calculated over a suitable range of values of ρ, for $n = 10$, 20 and 40, and taking $a = 1$ as a standard initial condition. Some typical results are shown in Figs. 1, 2 and 3. It can be seen from Figs. 1, 2 and 3 that when the relative removal rate ρ is large epidemics tend to be small, and

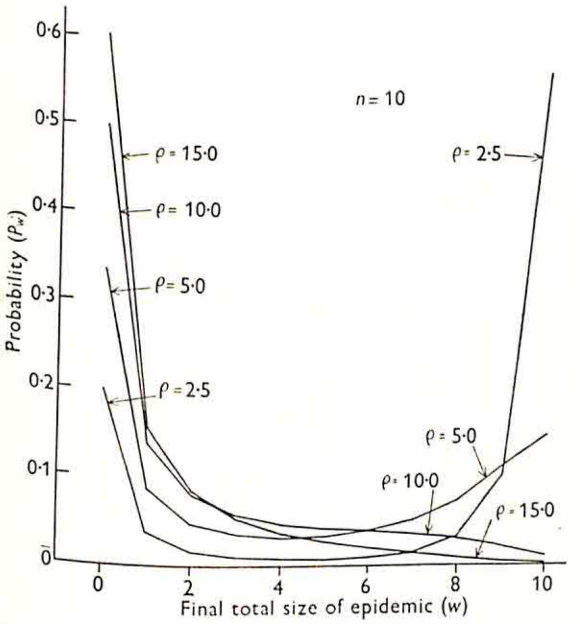


Fig. 1. Diagram showing the probability of the final total size of the epidemic for groups of ten susceptibles, starting with the introduction of one new infectious case.

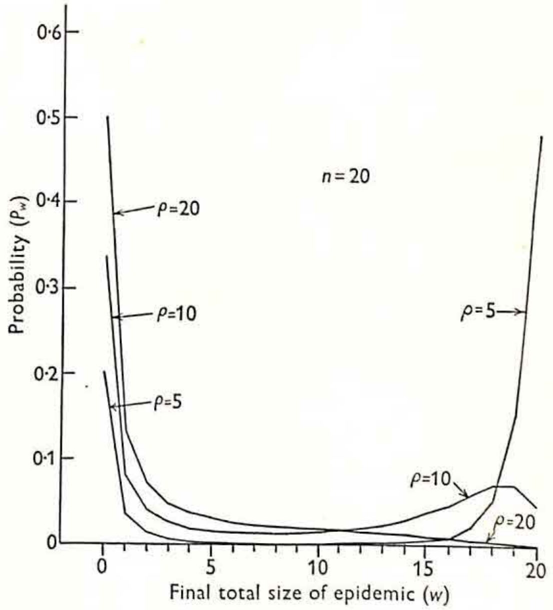


Fig. 2. Diagram showing the probability of the final total size of the epidemic for groups of twenty susceptibles, starting with the introduction of one new infectious case.

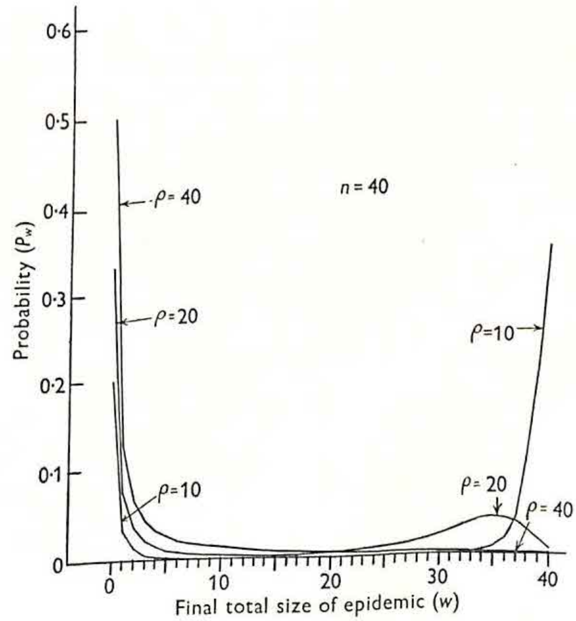


Fig. 3. Diagram showing the probability of the final total size of the epidemic for groups of forty susceptibles, starting with the introduction of one new infectious case.

conversely. There is a fairly gradual transition between the two extremes, though for some intermediate values of ρ most of the probability is accounted for by the two ends of the distribution. For example, with $\rho = 5$ for $n = 20$, there is a 20 % chance of no additional cases and a 64 % chance of 19 or 20. Again, there is only a gradual drop in the average size of the epidemic with increasing ρ. Specimen values for the range $\rho = 0\,(0{\cdot}25n)\,1{\cdot}50n$ are set out in Table 1 below. There is no obvious analogue of the Threshold Theorem derived by Kermack & McKendrick (1927) for the deterministic case.

Table 1. *Average total size of epidemic for various values of ρ and n (not counting initial case)*

ρ	$n = 10$	$n = 20$	$n = 40$
0	10·00	20·00	40·00
$0{\cdot}25n$	7·13	14·40	29·12
$0{\cdot}50n$	4·33	7·97	15·31
$0{\cdot}75n$	2·74	4·32	6·94
$1{\cdot}00n$	1·89	2·62	3·58
$1{\cdot}25n$	1·38	1·77	2·18
$1{\cdot}50n$	1·08	1·30	1·50

4. Household distribution of cases

The methods of the foregoing section may also be employed to investigate the distribution of multiple cases of a disease in households. This problem was first examined statistically by Greenwood (1931) who considered the hypothesis that, with a fairly infectious disease like measles, the first case in a family would arise from an outside contact while subsequent cases would occur through contacts within the family. The period of infectiousness is thought to be short for measles and if reduced, for the purpose of simplification, to an instant, Greenwood showed that the course of the intra-familial epidemic may be represented by a chain of binomial distributions. The frequencies of the final number of cases observed can then be found in terms of a parameter p, which is a measure of infectiousness. Such a model is quite adequate for measles, and satisfactory tests of goodness-of-fit were obtained for the data available, merely by equating observed and expected means. However, for diseases like diphtheria which have a more extended period of infectiousness there is probably some advantage in using the concepts of infection and removal employed in this paper. Equations (11), (14) and (15) can be used as before to calculate the quantities P_w for small values of n, e.g. 1 to 5, keeping in ρ as a parameter and taking a to be unity. There is some advantage, for simplicity in handling the algebra, in partially solving the recurrence relation in (15) to give g_{rs} as a linear function of $g_{r+1,i}$, $i = (s-1), \ldots, (n-r)$. The requisite formulae are easily found to be

$$\left.\begin{aligned} & g_{rs} = \sum_{i=s-1}^{n-r} (r+\rho)^{s-i-1}\, g_{r+1,i} \quad (s>1), \\ \text{with}\quad & g_{r1} = g_{r2}/(r+\rho), \\ \text{and}\quad & g_{n1} = 1. \end{aligned}\right\} \qquad (17)$$

Table 2

$n=1$: $P_0=\rho/(\rho+1)$ $\hat{\rho}=a_0/a_1$ $I_\rho=N/\rho(\rho+1)^2$

$P_1=1/(\rho+1)$

$n=2$:

$$\begin{aligned}
&P_0=\rho/(\rho+2)\\
&P_1=2\rho^2/(\rho+2)\,(\rho+1)^2\\
&P_2=2(2\rho+1)/(\rho+2)\,(\rho+1)^2\\
&\frac{dL}{d\rho}=\frac{a_0+2a_1}{\rho}+\frac{2a_2}{2\rho+1}-\frac{2(a_1+a_2)}{\rho+1}-\frac{N}{\rho+2}
\end{aligned}$$

$n=3$:

$$\begin{aligned}
&P_0=\rho/(\rho+3)\\
&P_1=3\rho^2/(\rho+3)\,(\rho+2)^2\\
&P_2=6\rho^3(2\rho+3)/(\rho+3)\,(\rho+2)^2\,(\rho+1)^3\\
&P_3=6(5\rho^3+12\rho^2+8\rho+2)/(\rho+3)\,(\rho+2)^2\,(\rho+1)^3\\
&\frac{dL}{d\rho}=\frac{a_0+2a_1+3a_2}{\rho}+\frac{2a_2}{2\rho+3}+\frac{(15\rho^2+24\rho+8)\,a_3}{5\rho^3+12\rho^2+8\rho+2}-\frac{3(a_2+a_3)}{\rho+1}-\frac{2(a_1+a_2+a_3)}{\rho+2}-\frac{N}{\rho+3}
\end{aligned}$$

$n=4$:

$$\begin{aligned}
&P_0=\rho/(\rho+4)\\
&P_1=4\rho^2/(\rho+4)\,(\rho+3)^2\\
&P_2=12\rho^3(2\rho+5)/(\rho+4)\,(\rho+3)^2\,(\rho+2)^3\\
&P_3=24\rho^4(5\rho^3+27\rho^2+47\rho+27)/(\rho+4)\,(\rho+3)^2\,(\rho+2)^3\,(\rho+1)^4\\
&P_4=24(14\rho^6+93\rho^5+235\rho^4+293\rho^3+197\rho^2+74\rho+12)/(\rho+4)\,(\rho+3)^2\,(\rho+2)^3\,(\rho+1)^4\\
&\frac{dL}{d\rho}=\frac{a_0+2a_1+3a_2+4a_3}{\rho}+\frac{2a_2}{2\rho+5}+\frac{(15\rho^2+54\rho+47)\,a_3}{5\rho^3+27\rho^2+47\rho+27}\\
&\qquad+\frac{(84\rho^5+465\rho^4+940\rho^3+879\rho^2+394\rho+74)\,a_4}{14\rho^6+93\rho^5+235\rho^4+293\rho^3+197\rho^2+74\rho+12}\\
&\qquad-\frac{4(a_3+a_4)}{\rho+1}-\frac{3(a_2+a_3+a_4)}{\rho+2}-\frac{2(a_1+a_2+a_3+a_4)}{\rho+3}-\frac{N}{\rho+4}
\end{aligned}$$

$n=5$:

$$\begin{aligned}
&P_0=\rho/(\rho+5)\\
&P_1=5\rho^2/(\rho+5)\,(\rho+4)^2\\
&P_2=20\rho^3(2\rho+7)/(\rho+5)\,(\rho+4)^2\,(\rho+3)^3\\
&P_3=60\rho^4(5\rho^3+42\rho^2+116\rho+106)/(\rho+5)\,(\rho+4)^2\,(\rho+3)^3\,(\rho+2)^4\\
&P_4=\frac{120\rho^5(14\rho^6+177\rho^5+910\rho^4+2443\rho^3+3626\rho^2+2836\rho+918)}{(\rho+5)\,(\rho+4)^2\,(\rho+3)^3\,(\rho+2)^4\,(\rho+1)^5}\\
&P_5=\frac{\begin{array}{r}120(42\rho^{10}+596\rho^9+3604\rho^8+12{,}240\rho^7+25{,}941\rho^6+36{,}144\rho^5\\ +34{,}061\rho^4+21{,}952\rho^3+9456\rho^2+2448\rho+288)\end{array}}{(\rho+5)\,(\rho+4)^2\,(\rho+3)^3\,(\rho+2)^4\,(\rho+1)^5}\\
&\frac{dL}{d\rho}=\frac{a_0+2a_1+3a_2+4a_3+5a_4}{\rho}+\frac{2a_2}{2\rho+7}+\frac{(15\rho^2+84\rho+116)\,a_3}{5\rho^3+42\rho^2+116\rho+106}\\
&\qquad+\frac{(84\rho^5+885\rho^4+3640\rho^3+7329\rho^2+7252\rho+2836)\,a_4}{14\rho^6+177\rho^5+910\rho^4+2443\rho^3+3626\rho^2+2836\rho+918}\\
&\qquad+\frac{\begin{array}{r}(420\rho^9+5364\rho^8+28{,}832\rho^7+85{,}680\rho^6+155{,}646\rho^5+180{,}720\rho^4\\ +136{,}244\rho^3+65{,}856\rho^2+18{,}912\rho+2448)\,a_5\end{array}}{\begin{array}{r}42\rho^{10}+596\rho^9+3604\rho^8+12{,}240\rho^7+25{,}941\rho^6+36{,}144\rho^5+34{,}061\rho^4\\ +21{,}952\rho^3+9456\rho^2+2448\rho+288\end{array}}\\
&\qquad-\frac{5(a_4+a_5)}{\rho+1}-\frac{4(a_3+a_4+a_5)}{\rho+2}-\frac{3(a_2+a_3+a_4+a_5)}{\rho+3}\\
&\qquad-\frac{2(a_1+a_2+a_3+a_4+a_5)}{\rho+4}-\frac{N}{\rho+5}
\end{aligned}$$

On the other hand, with values of n as small as 1 to 5, it is probably just as easy to derive the P_w straight from Foster's formula (16).

We cannot expect to be able to estimate β and γ separately as the asymptotic distribution of epidemic size for infinite time yields no information about the time scale. For this we should require data giving the time intervals between successive infections in families with two or more cases.

Having calculated the P_w for any given family size, we can throw the results into a form suitable for the maximum likelihood estimation of ρ. N is the total number of families of a given size; and a_w is the observed number of families with a total of w cases in addition to the first one. The case $n = 0$ is trivial. For $n = 1$ there are simple expressions for the amount of information as well as the maximum likelihood estimate, while for $n \geqslant 2$, the information functions become increasingly awkward to handle, and the simplest procedure is to use the well-known method of calculating the observed amount of information from sufficiently close values of the score. The values of the P_w and the corresponding score for $n = 1, 2, 3, 4$ and 5, are set out below in Table 2.

The values of P_w can be conveniently checked by ensuring that their sum for any given n is unity. Although the scores for the larger n contain some awkward looking polynomials in ρ, there is little difficulty in practice, with the aid of *Barlow's Tables* and a calculating machine, in computing the score at a few trial values of ρ for the purposes of inverse interpolation. However, if such methods were to be used at all extensively it would be worth while considering the construction of special tables. The scores are linear functions of the observations and the coefficients of these observational quantities could be tabulated over a wide range of values of ρ.

Suitable data for the application of the above methods do not seem to be available, or at any rate are not readily accessible, apart from the material on the 1926 St Pancras measles epidemic used by Greenwood (1931). As already mentioned, Greenwood found that the chain-binomial model gave a satisfactory fit and so we should hardly expect the present model to give a very adequate description for measles, though it might for other diseases. This is in fact the case. I have carried out the maximum likelihood estimation of ρ on Greenwood's data for families up to total size 5, and in no case is a satisfactory fit obtained. It is therefore hardly worth giving the details of the calculations. The epidemiological implications are, however, important, for it has thus now been shown that in the case of measles a satisfactory fit is to be obtained neither by the simple binomial distribution to be expected if the disease were not highly infectious within families (Greenwood, 1931) nor by the distribution expected when there is both infection and removal of the types discussed in the present paper. On the other hand, the chain binomial model used by Greenwood (1931), appropriate to very short periods of high infectivity, is adequate.

5. Summary and conclusions

An investigation has been made of the total size, i.e. for infinite time, of a general stochastic epidemic involving both infection and removal by recovery, death or isolation. For small homogeneously mixing groups no analogue has been found of the Threshold Theorem derived by Kermack & McKendrick (1927) for the deterministic case. In stochastic models wide variations in the size of an epidemic can occur purely by chance with fixed infection and removal rates. This may have important consequences for the interpretation of epidemiological data. An application to the problem of the distribution of multiple cases of disease in a household is also considered. It is shown how maximum likelihood estimates of the ratio of removal to infection rate can be obtained from suitable data, and the appropriate maximum likelihood scores are given for families up to a total size of 5 (not including the first case). The model under discussion is not suitable for diseases like measles involving short periods of high infectivity, but its adequacy for other infections requires to be tested.

I am indebted to Miss Eva Rowland for undertaking the computations, on which Figs. 1, 2 and 3 and Table 1 were based.

REFERENCES

BAILEY, NORMAN T. J. (1950). A simple stochastic epidemic. *Biometrika*, **37**, 193.

BARTLETT, M. S. (1946). *Stochastic Processes* (notes of a course given at the University of North Carolina, 1946).

BARTLETT, M. S. (1949). Some evolutionary stochastic processes. *J.R. Statist. Soc.* B, **11**, 211.

BROWNLEE, J. (1918). Certain aspects of the theory of epidemiology in special relation to plague. *Proc. Roy. Soc. Med.* (Sect. Epid. and State Med.), p. 85.

GREENWOOD, M. (1931). On the statistical measure of infectiousness. *J. Hyg., Camb.*, **31**, 336.

GREENWOOD, M. (1946). The statistical study of infectious diseases. *J.R. Statist. Soc.* **109**, 87.

KERMACK, W. O. & MCKENDRICK, A. G. (1927 and later). Contributions to the mathematical theory of epidemics. *Proc. Roy. Soc.* A, **115**, 700; **138**, 55; **141**, 94.

ROSS, R. (1916 and later). An application of the theory of probabilities to the study of *a priori* pathometry. *Proc. Roy. Soc.* A, **92**, 204; **93**, 212; **93**, 225.

SOPER, H. E. (1929). Interpretation of periodicity in disease-prevalence. *J.R. Statist. Soc.* **92**, 34.

EXPERIMENTAL EVIDENCE CONCERNING CONTAGIOUS DISTRIBUTIONS IN ECOLOGY

By D. A. EVANS

King's College, Newcastle upon Tyne

Introduction

In order to get evidence about the applicability of theoretical contagious distributions, as much data as possible on plant and insect populations have been collected together. The goodness of fit has been considered for three contagious distributions. They are the negative binomial, the Pólya-Aeppli, and the Neyman Type A. The general conclusions are that, on the whole, plant quadrat counts are fairly well fitted by the Neyman Type A distribution, while insect counts are fitted by the negative binomial distribution and not by the other two distributions. It is suggested that this difference is possibly due to a greater degree of competition and overcrowding in the case of plants. In Part I experimental evidence is given and discussed. Mathematical formulae and charts required are given in Part II.

Part I. Experimental evidence

1·1. *Description of data*

The plant quadrat counts and insect population counts analysed in this paper are arranged in four tables. Plant counts made at eight British localities are given in Table 1, while Table 2 refers to counts made on three types of American prairie. Some counts of insects and larvae are given in Table 3, and two counts of moth eggs are given in Table 4. The various sources from which these data have been collected are as follows:

(i) Counts numbered 1–24, Table 1 A, were made by Archibald (1948) and consist of samples from five maritime and two grassland communities. The type of community to which a count belongs is indicated in Table 1 A by a letter at the head of the count. The key to this letter code and the names of the species concerned are given in Table 1 B. The populations were sampled by means of a 500 sq.cm. quadrat split up into twenty-five smaller quadrats, each of 20 sq.cm. Frequency distributions were then made up from 100 or 500 contiguous 20 sq.cm. quadrats. Except where otherwise stated, plant species with less than an average of 0·6 of an individual per quadrat have been excluded from the present discussion. Counts with low means give poor discrimination between the three theoretical distributions.

(ii) The remaining six counts in Table 1 A, numbered 25–30, have been given by Barnes & Stanbury (1951). The observations were made upon uniform level expanses of virgin mica dam, several feet thick and some 5000 sq.yd. in area, the waste product of clay pits. Various sizes of quadrat were used, and the individual dams from which the six sets of counts were taken were classified by Barnes according to their moisture content, height and stage of colonization. A summary of this information is given in Table 1 C.

Barnes's data deal with the initial colonization of virgin area by immigration and the subsequent development of colonies by reproduction. With the exception of the willow,

Table 1A. *Frequencies for various plant species*

Number of individuals	1*a*	2*a*	3*a*	4*a*	5*a*	6*a*	7*a*	8*b*	9*b*	10*b*
0	4	12	15	39	57	82	60	26	—	45
1	3	8	17	23	6	4	18	36	—	25
2	8	9	28	17	12	4	19	27	3	11
3	13	13	18	13	5	3	2	10	2	4
4	11	6	9	4	5	1	—	—	4	8
5	9	8	12	4	5	2	—	—	8	3
6	8	11	—	—	7	1	1	1	6	1
7	10	7	—		1	1	—	—	1	1
8	3	8	1		—	—			13	—
9	3	7	—		1	1			10	2
10	8	3			1	1			5	—
11	3	4			—	—			8	
12	4	1							5	
13	4	1							8	
14	—	—							5	
15	3	—							3	
16	2	1							6	
17	1	—							6	
18	—	—							2	
19	—	1							1	
20	—	—							2	
21	—	—							1	
22	—	—							—	
23	2	—							1	
24	—	—							—	
25	—	—								
26	—	—								
27	—	—								
29	—	—								
30	1	—								
31	—	—								
34	—	—								
N	100	100	100	100	100	100	100	100	100	100
k_1	6·99	5·05	2·31	1·32	1·58	0·67	0·68	1·26	10·59	1·37
k_2	27·56	14·53	2·68	2·00	5·42	3·37	1·03	1·12	22·53	3·71
k_3	226·81	44·52	2·51	2·61	18·85	20·79	2·07	1·21	31·08	14·00
Method 1: $\hat{a}^{(1)}$	2·94	1·88	0·16	0·51	2·43	4·04	0·51	0·00	1·13	1·71
$T_{\text{P.}}$	+67	−15·7	−1·00	−1·27	−8·22	−4·37	+0·08	−0·05	−35·5	−0·39
S.E. $(T_{\text{P.}})$	54	19	1·2	1	7·5	9	0·5	0·35	30	4
$T_{\text{N.A.}}$	+97*	− 6·8	−0·97	−1·09	−3·56	+1·09	+0·17		−28·8	+1·61
S.E. $(T_{\text{N.A.}})$	44	17	1·2	1	5·4	4	0·45		30	3·2
Method 2: $\hat{a}^{(2)}_{\text{P.}}$				0·80	3·62	4·75	0·66			1·43
$U_{\text{P.}}$				−0·38	−1·88	−0·48	−0·10			+0·38
S.E. $(U_{\text{P.}})$				0·29	1·16	0·84	0·13			0·44
$\hat{a}^{(2)}_{\text{N.A.}}$	1·82	2·09		0·72	2·60	3·24	0·60			1·20
$U_{\text{N.A.}}$	+ 7·85*	− 1·05		−0·27	−0·27	+0·53	−0·06			+0·70*
S.E. $(U_{\text{N.A.}})$	3·0	2·0		0·25	0·64	0·34	0·1			0·34

Note. In all the tables an asterisk against the value of a test statistic indicates that it is greater in magnitude than twice its estimated standard error: two asterisks denote that it exceeds three times its estimated standard error.

Table 1 A (*cont.*)

Number of individuals	11*b*	12*b*	13*c*	14*c*	15*d*	16*e*	17*f*	18*f*	19*f*	20*g*	21*g*
0	75	75	1	274	88	—	181	237	354	52	18
1	5	10	15	71	101	—	118	110	58	32	3
2	9	3	27	58	101	—	97	78	41	13	7
3	6	2	42	36	84	1	54	30	25	3	9
4	—	4	77	20	54	2	32	23	11	—	12
5	3	2	77	12	30	3	9	8	3		8
6	—	2	89	10	16	2	5	6	4		8
7	1	1	57	7	16	6	3	3	3		3
8	1	—	48	6	7	7	1	—	—		9
9		—	24	3	2	6		4	—		5
10		—	14	—	1	8		1	1		1
11		1	16	2		10					4
12			9	1		4					—
13			3			5					4
14			1			2					1
15						7					2
16						7					2
17						2					2
18						5					1
19						2					1
20						3					—
21						2					
22						3					
23						2					
24						2					
25						—					
26						3					
27						1					
29						1					
30						2					
31						1					
34						1					
N	100	100	500	500	500	100	500	500	500	100	100
k_1	0·71	0·78	5·80	1·31	2·43	14·21	1·41	1·18	0·66	0·67	5·71
k_2	2·35	3·39	6·43	4·23	4·01	46·39	2·25	2·70	1·75	0·67	23·64
k_3	9·67	19·26	7·63	18·34	7·36	250·47	3·99	9·36	6·37	0·56	99·81
Method 1: $\hat{a}^{(1)}$	2·31	3·34	0·11	2·24	0·65	2·26	0·59	1·29	1·66	0·00	3·14
$T_{\text{P.}}$	−1·64	−2·39	−0·18	−1·54	−1·38	+30·41	−0·69	+0·68	−0·26	−0·11	−44
S.E. ($T_{\text{P.}}$)	3·5	7	1·8	2·7	1·3	95	0·6	1·1	0·89	0·13	44
$T_{\text{N.A.}}$	+0·25	+1·96	−0·14	+1·73	−0·86	+66·8	−0·44	+1·66	+0·64		−16
S.E. ($T_{\text{N.A.}}$)	2	4	1·8	1·8	1·1	83	0·5	0·9	0·6		36
Method 2: $\hat{a}^{(2)}_{\text{P.}}$	2·94	3·42		2·34			0·78	1·17	1·81		4·66
$U_{\text{P.}}$	−0·45	−0·06		−0·14			−0·26	+0·14	−0·10		−8·67
S.E. ($U_{\text{P.}}$)	0·50	0·64		0·29			0·14	0·15	0·13		4·95
$\hat{a}^{(2)}_{\text{N.A.}}$	2·19	2·49		1·82	0·71		0·70	1·00	1·47		3·19
$U_{\text{N.A.}}$	+0·08	+0·67*		+0·55*	−0·13		−0·15	+0·34**	+0·13		−0·30
S.E. ($U_{\text{N.A.}}$)	0·26	0·31		0·19	0·23		0·12	0·11	0·08		3

Table 1A (*cont.*)

Number of individuals	22*g*	23*g*	24*g*	25*h*	26*h*	27*h*	28*h*	29*h*	30*h*
0	12	53	65	101	124	41	150	66	63
1	22	32	18	25	76	14	34	19	19
2	19	10	10	26	33	5	28	9	11
3	17	2	5	36	17	2	8	2	6
4	15	3	2	18	2	1	4	2	1
5	6	—	—	21	—	—	2	1	—
6	5			6		—	2	1	
7	2			4		1	—	—	
8	2			6		—			
9	—			5					
10				3					
11				1					
N	100	100	100	252	252	64	228	100	100
k_1	2·61	0·70	0·61	2·22	0·80	0·64	0·67	0·62	0·63
k_2	3·69	0·90	0·99	6·55	0·93	1·44	1·33	1·29	0·94
k_3	5·06	1·41	1·64	19·98	0·97	5·31	3·28	3·64	1·33
Method 1: $\hat{a}^{(1)}$	0·42	0·28	0·62	1·95	0·16	1·25	1·00	1·08	0·50
$T_{\text{P.}}$	−1·48	+0·03	−0·45	−7·85	−0·24	+0·77	−0·38	−0·07	−0·47
S.E. ($T_{\text{P.}}$)	2	0·32	0·52	4·7	0·19	1·25	0·66	1·00	0·43
$T_{\text{N.A.}}$	−1·26	+0·06	−0·33	−3·64	−0·23	+1·27	−0·05	+0·29	−0·39
S.E. ($T_{\text{N.A.}}$)	2	0·3	0·48	4	0·19	1·20	0·5	0·7	0·4
Method 2: $\hat{a}^{(2)}_{\text{P.}}$		0·21	0·83	2·86		0·88	1·18	0·98	0·73
$U_{\text{P.}}$		+0·06	−0·13	−2·03*		+0·24	−0·12	+0·06	−0·15
S.E. ($U_{\text{P.}}$)		0·08	0·14	0·79		0·18	0·13	0·16	0·13
$\hat{a}^{(2)}_{\text{N.A.}}$	0·43	0·20	0·74	2·15	0·24	0·78	1·02	0·86	0·66
$U_{\text{N.A.}}$	−0·04	+0·06	−0·07	−0·44	−0·06	+0·30*	−0·01	+0·13	−0·10
S.E. ($U_{\text{N.A.}}$)	0·54	0·07	0·11	0·49	0·06	0·15	0·1	0·12	0·1

Table 1B. *List of plant species and localities referred to in Table* 1A

Plant Species

1*a* *Salicornia stricta*
2 *Plantago maritima*
3 *Limonium vulgare*
4 *Triglochin maritima*
5 *Armeria maritima*
6 *Festuca* sp.
7 *Suaeda maritima*
8*b* *Glaux maritima*
9 *Juncus gerardii*
10 *Aster tripolium*
11 *Triglochin maritima*
12 *Plantago maritima*
13*c* *Glaux maritima*
14 *Plantago coronopus*
15*d* *Carex arenaria*
16*e* *Salicornia stricta*
17*f* *Carex flacca*
18 *Briza media*
19 *Helictotrichon pratense*
20*g* *Thymus serpyllum*
21 *Festuca* sp.
22 *Bromus erectus*
23 *Carex* sp.
24 *Briza media*
25*h* *Hypnum schreberi*
26 *Spergularia rubra*
27 *Juncus effusus* seedlings
28 *Juncus effusus* seedlings
29 *Juncus* tussocks
30 *Juncus* tussocks

Localities

a Limonium Salt Marsh, Blakeney Point
b Juncus Marsh, Havant
c Glaux Low, Blakeney Point
d Carex Dune, Blakeney Point
e Salicornia Salt Marsh, Blakeney Point
f Chalk Grassland, Pitstone
g Chalk Grassland, Otford
h The Devonshire Moors

Table 1c. *Classification of the habitats sampled in locality h, Table* 1B, *together with the size of quadrat used*

25 L, D, H, 39 sq.cm.
26 I, D, H, dense, 39 sq.cm.
27 I, L, 156 sq.cm.
28 L, H, 600 sq.cm.
29 L, L, 2500 sq.cm.
30 L, H, 2500 sq.cm.

I or L Initial or late stage of colonization.
W or D Wet or dry.
H or L Highland or lowland areas.

Salix atrocinerea, the underlying vegetation was completely obliterated by the mica deposits. In view of the wind dispersal and the uniform surface one would expect a Poisson distribution for a quadrat count of willow seedlings, and this was in fact borne out by a sample of sixty-four quadrats of area 156 sq.cm., for which the mean number of seedlings per quadrat was 0·734. The water used to transport the mica in suspension was derived from moorland streams and probably carried in the separate protonemata of the mosses. Each protonema then gave rise to a number of upright shoots, and counts of the latter gave evidence of contagion, as shown in Table 1A, species numbered 25*h*. As soon as the dams dried out, the herb *Spergularia rubra* was found. The count given (26*h*) was taken on ground bare of parent plants. It seems likely that small aggregates of seeds were carried in by animals, which frequented the dams as soon as they were firm enough. The rush, *Juncus effusus*, was one of the earliest colonizers, and was also one of the most persistent. The count of *Juncus* seedlings (28*h*) was made on an older dam with numerous well-established but widely spaced clumps of rush, from which seed was possibly distributed by wind and by grazing animals.

(iii) Turning now to the plant quadrat counts made in America, Table 2A gives some results for two series of counts made by Steiger (1930) in Nebraska. Steiger gave the number of individuals per quadrat for all species found in two sets of forty quadrats. Each quadrat was 1 m. square, and those species with less than an average of one individual per quadrat have been excluded from the present discussion. The two sets of quadrats were placed in vegetation described as 'high-prairie' and 'low-prairie' respectively. The quadrats were selected at regular intervals along four parallel lines, and a detailed study of numerous small areas showed that the propagation of prairie species was largely vegetative. The results of calculations on counts for twenty high-prairie and eighteen low-prairie species are given in Table 2A. For both prairie types, the results are arranged in two groups according to Steiger's classification of the species into grasses and non-grasses. The species within each of the groups are arranged in descending order of density, and numbered accordingly. These numbers are also given in Table 2B, which gives the letter code, used by Steiger, for the names of the species.

Clapham (1936) has previously discussed Steiger's data, and he showed that, for the great majority of species, individuals do not follow the Poisson form of distribution.

(iv) The remaining American plant count is given in Table 2C. This count was made by Hanson (1934) on the Native Prairie of western North Dakota, and gives the number of stalks of the grass *Agropyron Smithii* per quadrat for each of 384 quadrats of area 0·1 sq.m. Sampling was by restricted randomization, as follows. Twelve large areas, each split into eight subplots, were laid out systematically over a fairly level plateau top. Each subplot was 5·5 yards square and four samples were taken from it by placing a wooden quadrat, 2·5 by

Table 2A. *Statistics calculated from frequency distributions for thirty-eight American plant counts, each based on 40 quadrats (Steiger's data)*

Grasses and sedges, high-prairie sample

Ranking species	1 *Af*	2 *Bc*	3 *As*	4 *Po*	5 *B*	6 *Cm*	7 *St*	8 *Sp*	9 *Pa*	10 *K*	11 *Cp*	12 *El*
n_0	1	0	0	0	8	2	7	9	2	2	7	33
k_1	38·9	38·4	28·4	25·3	17·8	16·2	12·8	12·1	10·6	10·7	5·03	1·27
k_2	908	681	502	950	828	280	264	134	71·4	50·8	33·5	19·95
k_3	14,869	12,854	7,340	61,112	43,868	4,338	7,506	1,077	575	256	248	379
$\hat{a}^{(1)}$	22·33	16·74	16·70	36·54	45·44	16·28	19·66	10·10	5·73	3·73	5·67	14·65
$T_{\text{N.B.}}$	−26,581	−10,638	−9,933	− 9,314	−32,189	−5,048	−3,134	−1,764	−315	−174	−165	−226
$T_{\text{P.}}$	−16,880	− 5,257	−5,977	+ 7,592	−13,787	−2,903	− 666	−1,148	−141	− 99	− 85	− 89
$T_{\text{N.A.}}$	− 7,179	+ 124	−2,021	+24,498	+ 4,615	− 758	+1,803	− 532	+ 33	− 25	− 4	+ 48
$E_{\text{P.}}(n_0)-n_0$	+0·6	+0·6	+1·8	+11	+11	+ 5	+ 5	−4	+0·6	−1·1	+3·8	+1
$E_{\text{N.A.}}(n_0)-n_0$	+6	+4	+7·3	+20	+19	+13	+14	+3	+4·3	+0·4	+9·5	+4

Grasses and sedges, low-prairie sample

Ranking species	1 *Af*	2 *Po*	3 *Bc*	4 *As*	5 *Cm*	6 *Pa*	7 *St*	8 *K*	9 *El*
n_0	0	1	10	3	8	4	7	13	25
k_1	111	86·4	24·5	22·28	6·97	6·73	5·40	3·30	2·62
k_2	1,575	829	761	466	50·08	50·10	26·81	12·06	37·98
k_3	37·7	−26,864	23,167	14,435	778·02	754·00	234·35	45·22	760·75
$\hat{a}^{(1)}$	13·12	8·60	30·06	19·91	6·18	6·45	3·96	2·66	13·47
$T_{\text{N.B.}}$	—	—	−23,356	−4,580	+109	+ 58	− 5·1	−31	−300
$T_{\text{P.}}$	—	—	−12,283	− 164	+242	+197	+37·4	−19	− 62
$T_{\text{N.A.}}$	−23,661	−35,654	− 1,211	+4,251	+375	+337	+79·8	− 8	+176
$E_{\text{P.}}(n_0)-n_0$	+0	−1	−1·3	+ 2·2	−0·7	+ 4	−0·5	−3·3	+3·5
$E_{\text{N.A.}}(n_0)-n_0$	+0	−1	+7·7	+10·1	+5	+10	+3·4	−0·4	+7·9

Other herbs, high-prairie sample

Ranking species	1 *a*	2 *At*	3 *S*	4 *P*	5 *E*	6 *H*	7 *Cs*	8 *Am*
n_0	0	20	21	6	15	20	18	24
k_1	9·88	6·02	3·23	2·68	2·17	1·85	1·37	1·35
k_2	52·37	134	62·95	7·10	6·66	10·28	5·98	5·87
k_3	228	4,251	2,318	29·80	23·55	82·28	56·58	32·54
$\hat{a}^{(1)}$	4·30	21·30	18·52	1·65	2·06	4·56	3·35	3·35
$T_{\text{N.B.}}$	−275	−1,605	− 76	−0·75	−10·59	−21·8	+10·48	−12·7
$T_{\text{P.}}$	−183	− 238	+ 477	+2·90	− 5·97	− 2·6	+18·21	− 5·1
$T_{\text{N.A.}}$	− 92	+1,128	+1,030	+6·55	− 1·34	+16·7	+25·93	+ 2·5
$E_{\text{P.}}(n_0)-n_0$	+2	+ 4	+ 8	+3	−1·2	+3	+6	+0·1
$E_{\text{N.A.}}(n_0)-n_0$	+4	+10	+13	+5	+0·9	+7	+9	+3·1

Non-grassy species, low-prairie sample

Ranking species	1 *At*	2 *Es*	3 *H*	4 *S*	5 *A*	6 *Am*	7 *Pr*	8 *Cs*	9 *E*
n_0	9	14	7	12	6	20	21	15	12
k_1	14·57	8·58	6·62	6·45	4·70	2·52	2·38	1·82	1·73
k_2	145·48	129·23	82·65	67·74	31·81	16·82	15·06	5·17	2·51
k_3	953·42	2,485	2,219	880·32	424·33	146·45	132·80	17·25	2·88
$\hat{a}^{(1)}$	8·98	14·07	11·48	9·50	5·77	5·66	5·34	1·83	0·45
$T_{\text{N.B.}}$	−1,805	−1,280	+ 240	−475	+ 26	−61	−43·1	−6·91	−1·93
$T_{\text{P.}}$	−1,218	− 431	+ 676	−184	+104	−20	− 9·2	−3·84	−1·75
$T_{\text{N.A.}}$	− 630	+ 417	+1,112	+108	+182	+20	+24·6	−0·77	−1·57
$E_{\text{P.}}(n_0)-n_0$	−6·2	−0·2	+ 8	+1	+ 5·9	+0·7	−0·1	+0·4	−2·2
$E_{\text{N.A.}}(n_0)-n_0$	−1·1	+8	+15	+8	+11·7	+5·6	+4·7	+2·3	−2·0
$\hat{a}^{(2)}_{\text{P.}}$	17·54	14·34	5·60	8·72	2·96	5·29	5·37	1·72	0·87
$U_{\text{P.}}$	−125	−2	+39	+5	+13	+1	−0·07	+0·2	−0·7
S.E. $(U_{\text{P.}})$	68	32	11	15	5	4	3·8	1·1	0·7
$\hat{a}^{(2)}_{\text{N.A.}}$	9·77	8·17	3·71	5·33	2·20	3·54	3·58	1·40	0·77
$U_{\text{N.A.}}$	−12	+51	+51	+27	+17	+5	+4·2	+0·79	−0·54
S.E. $(U_{\text{N.A.}})$	25	12	6	7	3	2	1·9	0·8	0·5

Table 2B. *Key to species referred to in Table 2A*

Grasses and sedges				Other herbs			
Prairie type		Species		Prairie type		Species	
Low	High			Low	High		
1	1	*Af*	*Andropogon furcatus*	5	1	*A*	*Amorpha canescens*
4	3	*As*	*A. scoparius*	6	8	*Am*	*Aster multiflorus*
—†	5	*B*	*Bouteloua hirsuta*	1	2	*At*	*Antennaria campestris*
			B. oligostachya	8	7	*Cs*	*Cathartolinum sulcatum*
3	2	*Bc*	*B. curtipendula*	9	5	*E*	*Erigeron ramosus*
5	6	*Cm*	*Carex meadii*	2	—	*Es*	*Equisetum laevigatum*
—	11	*Cp*	*C. pennsylvania*	3	6	*H*	*Helianthus scaberrimus*
9	12	*El*	*Elymus canadensis*	—	4	*P*	*Psoralea floribunda*
8	10	*K*	*Koeleria cristata*	7	—	*Pr*	*P. argophylla*
6	9	*Pa*	*Panicum scribnerianum*	4	3	*S*	*Solidago glaberrima*
			P. wilcoxianum				
2	4	*Po*	*Poa pratensis*				
—	8	*Sp*	*Sporobolus heterolepis*				
7	7	*St*	*Stipa spartea*				

† A dash is used to indicate either complete absence, or a density of less than one individual per quadrat.

Table 2C. *Frequency distribution of the number of stalks of* Agropyron Smithii (*Hanson's data*)

r	n_r	r	n_r	r	n_r	r	n_r	r	n_r
0	21	15	7	30	9	45	2	60	2
1	6	16	7	31	7	46	2	61	1
2	11	17	8	32	10	47	2	62	—
3	17	18	9	33	7	48	2	63	3
4	7	19	7	34	4	49	1	64	2
5	18	20	9	35	7	50	3	65	—
6	10	21	9	36	4	51	1	66	1
7	9	22	5	37	4	52	1	67	2
8	11	23	5	38	8	53	1	68	1
9	9	24	7	39	4	54	4	69	1
10	5	25	7	40	5	55	1	79	1
11	7	26	7	41	2	56	2	84	1
12	8	27	3	42	3	57	3	95	1
13	7	28	7	43	5	58	2	96	1
14	9	29	4	44	4	59	—	108	1

Statistics			
N	384	k_2	350·88
k_1	22·79	k_3	7324·96
$\hat{a}^{(1)}$	14·39		
Negative binominal:		$\hat{a}^{(2)}_{\text{N.B.}}$	25·79
$T_{\text{N.B.}}$	−3127	$U_{\text{N.B}}$	−260*
S.E. $(T_{\text{N.B.}})$	1830	S.E. $(U_{\text{N.B.}})$	77
$E_{\text{N.B.}}(n_0)-n_0$	−15·9		
Pólya-Aeppli:		$\hat{a}^{(2)}_{\text{P.}}$	13·69
$T_{\text{P.}}$	−766	$U_{\text{P.}}$	+16
S.E. $(T_{\text{P.}})$	1280	S.E. $(U_{\text{P.}})$	30
$E_{\text{P.}}(n_0)-n_0$	+2·8		
Neyman A:		$\hat{a}^{(2)}_{\text{N.A.}}$	7·84
$T_{\text{N.A.}}$	+1596	$U_{\text{N.A.}}$	+149*
S.E. $(T_{\text{N.A.}})$	952	S.E. $(U_{\text{N.A.}})$	14
$E_{\text{N.A.}}(n_0)-n_0$	+57·8		

Table 3 A. *Frequency distribution for eleven insect population counts (data given by Beall)*

Number of individuals	1*a*	2*a*	3*a*	4*a*	5*b*	6*b*	7*b*	8*c*	9*c*	10*c*	11*c*
0	19	24	43	47	190	20	33	117	205	162	227
1	12	16	35	23	264	11	12	87	84	88	70
2	18	16	17	27	304	6	5	50	30	45	21
3	18	18	11	9	260	6	6	38	4	23	6
4	11	15	5	7	294	6	5	21	2	5	1
5	12	9	4	3	219	1	—	7	—	2	—
6	7	6	1	1	183	1	2	2		—	
7	8	5	2	1	150	2	2	2			
8	4	3	2	—	104	3	2	—			
9	4	4	—	—	90	—	—	1			
10	1	3		1	60	1	1	—			
11	—	—		1	46	2	—				
12	1	1		—	29	1	—				
13	1	—			36	1	—				
14	—				19	—	—				
15	1				12	1	—				
16	—				11	1	—				
17	1				6	—	1				
18	—				10	2	—				
19	1				2	—	—				
20	—				4	—	—				
21	—				1	1	1				
22	—				3	—	—				
23	—				4	—					
24	—				1	—					
25	—				1	—					
26	1				—	—					
27	—				—	1					
28					1	—					
36					—	1					
45						1					
60						1					
N	120	120	120	120	2304	70	70	325	325	325	325
k_1	4·03	3·17	1·48	1·51	4·74	6·10	2·14	1·40	0·50	0·85	0·41
k_2	16·45	7·77	3·19	3·63	15·00	113·13	14·10	2·33	0·58	1·14	0·52
k_3	155·11	19·75	9·67	15·63	83·83	3787·50	162·75	4·72	0·72	1·54	0·72

Table 3 B. *List of insect species and treatments referred to in Table 3 A*

Insect species:		
	a	The European corn-borer, *Pyrausta nubilalis* Hubn.
	b	The Colorado potato beetle, *Leptinotarsa decemlineata* Say.
	c	The beet webworm, *Loxostege sticticalis* L.

Treatments. Counts 1–4 were made on plots given two applications of fungus spores, at the following levels in grams per acre.

Count	Application on 8 July	Application on 19 July
1	0	0
2	0	40
3	40	0
4	40	40

Count 5 was made on a potato field near Chatham, Ontario, on 14 August 1935. No information is given about the treatments given to the beetles in counts 6 and 7. Counts 8–11 were on plots given various chemical treatments, as follows:

8	No spray	10	Lead arsenate
9	Contact insecticide	11	Contact spray and lead arsenate

4·0 dm., in each quarter. The vegetation is described as 'mixed prairie', and *A. Smithii* is one of the two dominant tall grasses found on the plateau.

This concludes the description of the plant quadrat counts.

(v) Passing on to the counts of insect and larvae populations, counts numbered 1–4 and 6–11 in Table 3A have been given by Beall (1940). One of the series of counts had to be excluded because it was not given in full. Table 3B gives the key to the species letter code, and the key to the number code for the treatments applied to the insect populations.

Counts 1–4 were made in an experiment (Stirrett, Beall & Timonin, 1937) on the control of the European corn-borer by the fungus *Beauveria Bassiana* Vuill. The fungus treatment on 8 July occurred at the beginning of the period of oviposition and the second treatment at its height. The borer lays masses of about twenty eggs, which hatch in July and reach full growth in August, so that the population sampled was effectively of one age. Counts were made of the number of borers on the unit area occupied by a hill of corn, and there had been two possible periods of migration of the larvae prior to the examination on 19 October.

Table 4. *Frequency distributions for two counts of the number of moth eggs per maize plant* (*Marshall's data*)

Number (r) of eggs per plant	Frequencies (n_r)		r	n_r 2nd count continued	Statistics		
	1st count	2nd count			Counts	1st	2nd
0	204	101	19	1	N	780	782
1	143	72	20	3	k_1	2·48	5·75
2	128	89	21	1	k_2	7·40	32·10
3	107	84	22	2	k_3	39·43	383·55
4	71	66	23	2		Method 1	
5	36	40	24	1	$\hat{a}^{(1)}$	1·98	4·58
6	32	56	25	—	$T_{\text{N.B.}}$	+ 2·67	+ 57·25
7	17	43	26	1	S.E. ($T_{\text{N.B.}}$)	4·4	40·4
8	14	49	27	1	$T_{\text{P.}}$	+ 7·55*	+ 117·62*
9	7	28	28	—	S.E. ($T_{\text{P.}}$)	3·3	28
10	7	39	29	2			
11	2	13	30	—		Method 2	
12	3	17	31	—	$\hat{a}^{(2)}_{\text{N.B.}}$	2·08	5·07
13	3	12	32	1	$U_{\text{N.B.}}$	− 0·23	− 2·8
14	1	14	33	—	S.E. ($U_{\text{N.B.}}$)	0·49	2·6
15	1	12	34	3	$\chi^2_{\text{N.B.}}$	11·9	40·3*
16	1	10	—	—	D.F.	9	18
17	2	11	51	1	$\chi^2_{\text{P.}}$	9·5	34·2*
18	1	7	—	—	$\hat{a}^{(2)}_{\text{P.}}$	1·70	3·62
					$U_{\text{P.}}$	+ 0·71*	+ 5·51*
					S.E. ($U_{\text{P.}}$)	0·34	1·5

The Colorado potato beetle lays groups of 20–30 eggs, and the two counts 6*b* and 7*b* were made on the number of larvae per 4 ft. strip of potato row. The population sampled ranged from larvae newly hatched to completely mature larvae.

The beet webworm lays from one to five eggs, and the larvae usually mature on the plant on which they have hatched, although they can move about freely. Counts 8*c*–11*c* were made on the number of larvae present on unit areas of 3 ft. of row, and the population consisted mainly of mature larvae with some half-grown larvae. Beall suggests that the contact insecticide, acting as an irritant, might possibly increase dispersion, whilst the arsenate might be expected to depress it.

(vi) Beall (1938) has also given an account of some earlier work on the Colorado potato beetle. A plot of potato plants forty-eight rows wide and 96 ft. long was chosen for examination. By running strings transversely to the rows of potatoes at intervals of 2 ft., the area was split up into 2304 sampling units of 2 ft. lengths of row. The frequency distribution of the number of beetles per unit is given in Table 3A, number 5*b*.

(vii) Finally, Table 4 gives two frequency distributions obtained in a field sampling study, by Marshall (1936), of oviposition by the moth, *Heliothis obsoleta* Fabr. Two counts of the number of eggs per maize plant were made for all the maize plants in a plot 23 by 24 yards. All the eggs found during the first count were destroyed, so that the second count, which took place a week later, was independent of the first count.

1·2. *Analysis of data*

The purpose of the present analysis is to see whether one can choose between various theoretical contagious distributions and say that, for a particular type of data, one form of distribution gives a better fit in general than the others. No claim is made that the typical form of distribution will always give a good fit. Individual cases will certainly arise where the fit is poor, and without further data it is impossible to say whether poor fits are a persistent feature for counts of certain species.

Because of the poor discrimination between alternative theoretical distributions usually shown by plant and insect population counts, it seems necessary to examine as a whole a large number of counts of each type, as many as one can get in fact, with no personal bias in the selection. The problem is then to detect whether one theoretical distribution always tends to err in a particular direction, whilst another gives a satisfactory median fit.

The Neyman Type A, Pólya-Aeppli and negative binomial distributions have been fitted according to the methods given below in Part II. These three distributions were chosen as being typical of the various forms of contagious distribution that have been proposed by various writers (see Anscombe, 1950). They are in increasing order of skewness and in decreasing order of proportion of zeros, for given mean and variance.

In order to test the relative adequacy of fit of each type of distribution, a statistic T or U has been calculated for each (sometimes both statistics were calculated). The statistic T is the difference between the sample estimate of the third cumulant and its expected value found by using the sample estimates of the first two moments; U is the difference between the sample estimate of the variance and its expected value found by using the sample estimates of the mean and proportion of zeros. A large positive value of either statistic suggests that a more skew form of distribution would give a better fit, while a large negative value suggests a less skew form of distribution. The relative sensitivity of the two statistics in testing goodness of fit depends on the values of the parameters of the parent distribution. The statistic T is relatively easy to calculate so that its value has been given throughout, and the value of U has been given whenever it is expected to be the more sensitive statistic. In general when the mean is small, for a given sample size, no appreciable discrimination between alternative forms of distribution is possible; and therefore counts with small means have been omitted, as stated above in the description of the data.

The general conclusion which appears from the application of these tests is that, on the whole, plant population counts are fitted by a less skew form of distribution than that which fits insect population counts. The explanation of this difference may be that in some manner the two types of population are differently affected by overcrowding and by competition. For the plant and insect populations considered here, the plants are larger in size on the average than the insects, and they may therefore be more prone to suffer from overcrowding, in the sense that their physical size prevents high densities per unit area from occurring. Competition by other species may be relatively slight in the case of insect communities on plant hosts if, as seems likely from the relevant literature, only a small number of insect species are able to tolerate one kind of plant host.

The population counts discussed here do not show any indication that a more skew form of distribution than the negative binomial distribution would give a better fit. However, in the course of searching the literature for insect population counts, four series of insect parasites on animal hosts were found, which showed a greater degree of skewness than could be fitted by this distribution. It may be possible to discuss these counts in a later paper, but so far no more excessively skew series of this type have been found.

Referring back to the thirty plant counts given in Table 1 A, it will be seen that only a few of the values of the statistics T and U (marked with an asterisk) differ from zero by more than twice their estimated standard error. Taking the group as a whole, the Neyman Type A distribution (N.A.) gives a satisfactory median fit and the Pólya-Aeppli distribution (P.) gives a significantly high proportion of negative values. Values of T and U for the negative binomial distribution (N.B.) are not given, but they also show a significantly high proportion of negative values.

In order to investigate the general spread of the statistics T and U for the Neyman Type A and Pólya-Aeppli distributions, the value of the more efficient statistic for each count was divided by its standard error and converted by a table of probits into a probability value. This use of a probit conversion is only strictly valid if T and U are normally distributed and the standard error precisely known. One would expect, however, that the error committed is not serious. It may be as well to stress here that all the theoretical results used in this analysis are based on large-sample theory. On the null hypothesis that the distribution fitted is the correct one, a rectangular distribution in $(0, 1)$ would be expected for the probability values. After dividing the interval into five equal parts, this hypothesis was tested by the χ^2 goodness of fit test. The values of χ^2 (with 4 degrees of freedom) were 5·7 and 12 for the Neyman Type A and Pólya-Aeppli distributions respectively, so that the latter distribution does not give a satisfactory fit.

The conclusion, that the Neyman Type A distribution gives a satisfactory median fit, may be verified by calculating and comparing expected and observed group frequencies, and this has been done in those cases where there are a large number of observations. The results are given in Table 5, using the same key as in Table 1 A. The Neyman Type A distribution was fitted by using the sample mean and proportion of zeros (Method 2, Part II below), except for count 13c, where the fitting was by moments (Method 1). Expected frequencies for the Pólya-Aeppli distribution fitted by Method 2 are also given for count 14c.

The value of χ^2 for the goodness of fit test and its number of degrees of freedom are given at the foot of each count. The total of 53·8 for χ^2 with 42 degress of freedom is not significantly large, nor is there any consistent failure in fitting at certain points, so that the conclusion given above is not discredited.

It may be noted before passing on that for count 14c, the χ^2 test is sufficiently sensitive to discriminate between the two forms of distribution—the result agreeing with the test based on the U statistic. The U test is also able to discriminate between the two forms of distribution for the count $18f$, but here, as is more usual, the χ^2 test fails to do so.

Table 5. *Expected frequencies for the Neyman Type A distribution fitted to some plant population counts from Table* 1

Number of individuals	13*c*		14*c*			15*d*		17*f*	
	Obs.	Exp.	Obs.	N.A.	P.A.	Obs.	Exp.	Obs.	Exp.
0	16	12·8	274	274·0	274·0	88	88·0	181	181·0
1			71	58·0	75·9	101	105·0	118	127·1
2	27	28·4	58	58·9	51·5	101	99·9	97	89·0
3	42	51·3	36	43·6	34·4	84	78·2	54	52·0
4	77	70·7	20	27·6	22·7	54	53·9	32	27·3
5	77	79·5	12	16·4	14·9	30	33·8	9	13·2
6	89	75·8	10	9·5	9·6	16	19·7	9	10·4
7	57	63·1	7	5·4	6·2	16	10·8		
8	48	46·7	12	6·4	10·8	7	5·6		
9	24	31·2				3	5·2		
10	14	19·1							
11	16	10·8							
12 and over	13	10·7							
N	500		500			500		500	
χ^2		12·2		12·8	2·4		5·5		3·8
D.F.		9		6	6		7		4

Number of individuals	18*f*		19*f*		25*h*		26*h*		28*h*	
	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
0	237	237·0	354	354·0	101	101·0	124	124·0	150	150·0
1	110	102·9	58	53·8	25	26·2	76	77·8	34	36·2
2	78	73·8	41	43·5	26	31·6	33	33·7	28	22·8
3	30	42·8	25	25·5	36	27·8	19	16·4	8	11·0
4	23	22·6	11	12·6	18	20·9			8	8·0
5	8	11·3	11	10·6	21	14·8				
6	14	9·7			6	10·3				
7					10	11·6				
8										
9					9	7·8				
10										
11										
12 and over										
N	500		500		252		252		228	
χ^2		7·5		0·7		8·6		0·5		2·2
D.F.		4		3		6		1		2

Note. The theoretical distributions fitted are all Neyman Type A except in the case of count $14c$ to which a Pólya-Aeppli distribution was also fitted.

None of the three contagious distributions considered here gives a satisfactory fit to the thirty-eight plant species referred to in Table 2A. When the fitting is done by Method 1, the Neyman Type A distribution gives the best fit as far as third moments are concerned, but goes wrong at the lower end of the frequency distribution, predicting too many zeros in thirty-four cases out of thirty-eight.* The negative binomial distribution would predict about the right proportion of zeros, but would generally be far too skew.

These conflicting results may be due to a number of causes. The communities sampled by Archibald were well-known natural forms of a more or less stable character and were chosen for their uniformity of life form. There were therefore no difficulties of sampling due to widely differing life forms. The degree of cover is more or less complete in the chalk grasslands, and especially in the *Salicornia* and *Limonium* marshes, where the vegetation forms a low compact mat. The structure of the prairie vegetation showed great variation, with several species at times locally dominant. The character of this community is entirely different from that of the communities studied by Archibald. Many of the grasses, the Andropogons and Boutelouas in particular, are sod (that is, tussock) forming with foliage from 30 to 65 cm. high. The formation of dense sods of *Bouteloua* is encouraged by the annual mowing of the area. *Andropogon furcatus* has a rank growth with widely spreading tops producing almost complete cover, yet it may occupy only 20 % of the surface area. In the case of the sod-forming species it seems certain that Steiger has given the number of flowering stems per quadrat, but he is not definite about this.

No experimental field work seems to have been published about the effect on the shape of the frequency distribution produced by using samples based on different quadrat sizes. Consequently it is not easy to judge the significance of the fact that Steiger's quadrat size is 500 times as large as the 20 sq.cm. quadrat used by Archibald. From the mathematical point of view, the three distributions considered here are only additive with respect to the parameter m; that is, if two independent random variables have the same type of distribution, but with parameters (m_1, a) and (m_2, a) respectively, then their sum also has the same type of distribution, but with parameters (m_1+m_2, a).

Although the quadrat size was so much larger, Steiger only found from ten to twenty stems per square metre for each of the more abundant species, and about 300 stems for all species combined. The corresponding figures obtained by Archibald were one to two and fifteen, so that the English communities had about twenty-five times as many individuals per unit area, with fewer species having about the same, high, average number of individuals per unit area.

It is possible that the Neyman Type A distribution is the correct form of distribution, and that the annual mowing and the higher number of equal competitors have spread out the individuals of each species more evenly than usual, thereby reducing the number of zeros observed.

An alternative conjecture is that the negative binomial distribution is the appropriate distribution, and that the upper tail has been compressed because any high densities are physically unlikely for large, rank-growing, species.

If n_0 is the observed number of zeros and $E(n_0)$ is its estimated expected value obtained by fitting the distribution by moments, then $E(n_0)-n_0 = V$ say, and $U = k_2 - k_1(1+\hat{a}^{(2)})$

---

* I.e. thirty-four out of the thirty-eight values of $E_{\text{N.A.}}(n_0)-n_0$ shown at the bottom of the table are positive.

have the same sign for any of the three distributions considered here; and further, since $U_{\text{N.A.}} > U_{\text{P.}} > U_{\text{N.B.}}$, ($\hat{a}^{(2)} > 0$), it follows that

$$E_{\text{N.A.}}(n_0) - n_0 > E_{\text{P.}}(n_0) - n_0 > E_{\text{N.B.}}(n_0) - n_0.$$

This means that the statistics U and V can be regarded as equivalent as far as the study of the proportion of positive to negative signs is concerned. Thus if thirty-four positive values and four negative values of V are found for the Neyman Type A distribution, thirty-four out of the thirty-eight values of U would also be positive.

The statistics $U_{\text{P.}}$ and $U_{\text{N.A.}}$ are given for comparison for the nine low-prairie, non-grassy species. Their estimated standard errors have also been given, although for N as low as 40 they will not be very reliable.

Table 6. *The Neyman Type A distribution fitted to five high-prairie, non-grassy species from Table* 2

Number of individuals	4*P*		5*E*		6*H*		7*Cs*		8*Am*	
	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
0	6	6·0	15	15·0	20	20·0	18	18·0	24	24·0
1	12	7·7	12	10·5	11	7·4	10	7·4	5	2·9
2	7	7·8					6	6·0	4	3·7
3	7	11·2	7	7·9	4	6·8	6	8·6	1	3·3
4									1	2·3
5 and over	8	7·3	6	6·6	5	5·8			5	3·8
χ^2	4·1		0·4		2·9		1·7		—	
D.F.	2		1		1		1		—	
$U_{\text{N.A.}}$	+2·5		+0·4		+3·9		+2·9		+1·3	

The Neyman Type A distribution has been fitted to five high-prairie, non-grassy, species by Method 2. The results are given in Table 6, together with the values of χ^2 and the statistic $U_{\text{N.A.}}$. There are not enough degrees of freedom to apply a χ^2 test to the fit for count 8*Am*, and the total χ^2 of 9·1 on 5 degrees of freedom for the remaining counts is not significantly large. Inspection of individual frequencies confirms the picture given by the U statistics.

The final plant count given in Table 2c is also for a tall prairie grass. Gray's *Manual of Botany* gives 3–15 dm. as the height of the stems of *Agropyron Smithii*. The area was sampled with an 0·1 sq.m. quadrat, and was selected according to the following criteria: 'typical and homogeneous vegetation, uniform topography and soil, freedom from erosion, mowing and grazing or other disturbing factors.' The Pólya-Aeppli distribution gives the best fit. The T and U statistics both indicate that a better fit would be given by a more skew form of distribution than the Neyman Type A.

The analysis of the insect counts given in Table 3A will now be considered. The negative binomial distribution gives a satisfactory fit to these counts.

When Beall tried to fit this distribution and the Pólya-Aeppli distribution (which he calls

Pólya Types 1 and 2 respectively), he got his parameters interchanged and was consequently unfair to them. Furthermore, his method of estimation is always by the first two moments, and this method is seen to be particularly inefficient in the case of the Colorado potato beetle counts (5*b*, 6*b* and 7*b*).

In addition to the T and U statistics, extensive calculations of expected values were carried out for the distributions in Table 3 A. Expected values were calculated for the negative binomial and Pólya-Aeppli distributions, fitted by both methods, but restrictions of space prevent the presentation of this information in full. Table 7 A gives the values of the fitted parameters and of χ^2 for both methods of fitting, for both distributions. None of the values of χ^2 is significant for the negative binomial distribution, both methods of fitting giving reasonable values of χ^2. The total χ^2 of 55 for the preferred method of fitting has 50 degrees of freedom and is not significantly large. Two values of χ^2 are significant for the Pólya-Aeppli distribution (counts 5*b* and 6*b*), and the total χ^2 of 77 on 50 degrees of freedom is significant at 1 %. If, however, the large significant χ^2 for count 5*b* is removed, the total χ^2 is found to be insignificant at 36 on 32 degrees of freedom. A significant value of χ^2 produced by a poor method of fitting only occurs once, in the case of count 7*b*.

Table 7 A. *Statistics for the negative binomial and Pólya-Aeppli distributions fitted by Methods* 1 *and* 2 *to the insect population counts of Table* 3

Count N	1*a* 120	2*a* 120	3*a* 120	4*a* 120	5*b* 2304	6*b* 70	7*b* 70	8*c* 325	9*c* 325	10*c* 325	11*c* 325
$\hat{a}^{(1)}$	3·08	† 1·45	1·15	1·41	† 2·17	17·55	5·58	† 0·66	† 0·16	† 0·34	0·26
$T_{\text{N.B.}}$	+37·3	−10·6	−0·9	+1·8	+3·8	−296	−8·6	−0·69	−0·05	−0·37	−0·07
S.E. (T)	36	8·8	3·0	4·1	5·5	1200	89	0·89	0·09	0·28	0·11
χ^2	6·8	6·2	1·2	7·0	21·1	3·7	0·12	5·2	1·43	4·2	0·33
D.F.	7	5	3	3	18	4	2	4	1	2	1
$\hat{a}^{(2)}_{\text{N.B.}}$	3·07	2·42	1·01	1·43	2·22	12·77	5·20	0·82	0·20	0·48	0·31
$U_{\text{N.B.}}$	+0·0	−3·1	+0·22	−0·03	−0·25	+29·1	+0·8	−0·22	−0·02	−0·12	−0·02
S.E. (U)	2·7	1·8	0·41	0·56	0·69	25	3·7	0·20	0·03	0·08	0·03
χ^2	6·8	4·9	1·3	7·1	21·6	1·4	0·08	5·3	1·51	3·7	0·11
D.F.	7	5	3	3	18	4	2	4	1	2	1
$T_{\text{P.}}$	+56·5*	−7·3	+0·1	+3·3	† +14·9*	+643	24·7	−0·38	−0·04	−0·32	−0·06
S.E. (T)	27	7·4	2·3	3·1	4·4	1060	49	0·75	0·09	0·26	0·10
χ^2	8·6	3·6	2·8	6·9	41·2*	50·5*	8·1*	3·8	1·30	3·5	0·20
D.F.	7	5	3	3	18	3	2	4	1	2	1
$\hat{a}^{(2)}_{\text{P.}}$	2·38	1·94	0·89	1·22	1·80	7·74	3·70	0·74	0·20	0·45	0·30
$U_{\text{P.}}$	+2·8	−1·5	+0·39	+0·28	+1·76*	+59·8*	+4·0*	−0·11	−0·02	−0·10	−0·01
S.E. (U)	1·9	1·2	0·32	0·39	0·48	5·9	1·9	0·16	0·03	0·07	0·03
χ^2	5·0	2·4	2·2	5·7	45·9*	10·3*	2·0	3·7	1·34	2·9	0·04
D.F.	7	5	3	3	18	4	2	4	1	2	1

Note. For the cases marked †, Method 1 is expected to be the more efficient method of fitting; for the remaining cases Method 2 is preferable.

The fits by the preferred method to counts 5*b* and 6*b* for both distributions have been given in Table 7 B. The negative binomial distribution gave similar results for the other insect population counts. This table also illustrates a persistent feature for the counts as a whole, which is that an observed cell frequency is usually either greater than, or less than, the expected frequency for both types of distribution. Only very rarely is the observed frequency greater than one expectation and less than the other.

The negative binomial distribution gives a satisfactory fit to Marshall's first moth egg count, given in Table 4. The values of T and U are both greater than twice their estimated standard errors for the Pólya-Aeppli distribution, although χ^2 for the distribution fitted by Method 2 is only just greater than expectation.

Table 7B. *The negative binomial* (N.B.) *and Pólya-Aeppli* (P.) *distributions fitted to two insect population counts*

Count 5b				Count 5b continued				Count 6b			
r	Obs.	N.B.	P.	r	Obs.	N.B.	P.	r	Obs.	N.B.	P.
0	190	185·5	237·3	12	29	37·7	40·0	0	20	20·0	20·0
1	264	277·3	258·8	13	36	28·2	29·5	1	11	8·9	5·1
2	304	302·3	275·8	14	19	20·9	21·5	2 3	12	10·7	9·1
3	260	288·6	268·2	15	12	15·4	15·5				
4	294	256·0	245·1	16	11	11·3	11·1	4 5 6	8	9·5	10·8
5	219	216·7	213·9	17	6	8·3	7·9				
6	183	177·6	180·1	18	10	6·1	5·6				
7	150	142·1	147·3	19 20	6	7·6	6·6	7–10	6	7·4	10·0
8	104	111·6	117·6					11–18	8	7·4	10·0
9	90	86·4	92·0	over 20	11	8·2	5·8	over 18	5	6·1	4·9
10	60	66·2	70·7								
11	46	50·2	53·5	χ^2_{18}		21·1	41·2*	χ^2_4		1·4	10·3*

The negative binomial distribution is possibly the appropriate distribution for Marshall's second count. Both T and U are significantly positive for the Pólya-Aeppli distribution, and the values of these statistics are insignificant for the former distribution. Expected frequencies were found for the two distributions fitted by Method 2, and both gave significantly large values of χ^2. Comparing the observed with the expected series, there is a suggestion that even values of r have been favoured at the expense of odd values. The negative binomial distribution was also fitted by maximum likelihood, giving a value of 36·5 for χ^2 with 18 degrees of freedom. The same feature was again suggested; two-thirds of the value of χ^2 being contributed by large positive differences, observation − expectation, for cells 0, 8 and 10, and large negative values for cells 1 and 5. Marshall mentions that the 'zero' and 'few' egg frequencies may be in error because of the minuteness of the eggs.

PART II. MATHEMATICAL FORMULAE

2·1. *Introduction*

The application of two methods of fitting and the corresponding tests of departure from the fitted form of distribution are considered for each of the three contagious distributions in turn.

The two methods of fitting are: *Method* 1, the use of the mean and variance, and *Method* 2 the use of the mean and proportion of zeros. The usefulness of the latter method as an alternative to the method of moments for the negative binomial distribution has been pointed out by Anscombe (1950), who also gave the criteria, T and U, of the tests of departure, with particular reference to the negative binomial distribution. These criteria have been described above in the first part of the section on the analysis of the data (p. 196).

Some results for the Poisson limiting form of the three distributions are given after the following section on notation.

2·2. *Notation*

N	is the total number of observations in the sample,
n_r	is the number of observations equal to r $(r \geqslant 0)$,
k_1, k_2, k_3	are the first three k-statistics,
P_r	is the probability of observing any non-negative integer r,
$\kappa_i, \kappa_{[i]}$	denote the ith cumulant and the ith factorial cumulant, and are defined by

$$\ln E\{e^{tr}\} \equiv \sum_{i=1}^{\infty} \frac{\kappa_i t^i}{i!}, \quad \ln E\{(1+t)^r\} \equiv \sum_{i=1}^{\infty} \frac{\kappa_{[i]} t^i}{i!}.$$

The estimates of a parameter by Methods 1 and 2 are distinguished by attaching an appropriate upper suffix, in brackets. For example, $\hat{a}^{(1)}$ and $\hat{a}^{(2)}$ denote the sample estimates of the parameter a by Methods 1 and 2 respectively. It may be noted here that both $\hat{a}^{(1)}$ and $\hat{a}^{(2)}$ should be non-negative. If either of the estimates is found to be negative, the value zero is arbitrarily assigned to it, implying that a Poisson distribution is being fitted. To distinguish the three distributions, negative binomial, Pólya-Aeppli and Neyman Type A, suffix initials, for example $\hat{a}^{(1)}_{\text{N.B.}}$, $\hat{a}^{(1)}_{\text{P.}}$, and $\hat{a}^{(1)}_{\text{N.A.}}$, are used when this is required.

2·3. *Poisson limiting form of distribution*

All three distributions have mean m and variance $m(1+a)$, and they all tend to the Poisson form as $a \to 0$. The values of the variances of the various estimates and statistics have the same limiting values for all three distributions, and these are given below. The large-sample variances and covariances have been obtained by treating $\hat{m}^{(1)} - m$, $\hat{a}^{(1)} - a$, and $\hat{a}^{(2)} - a$ as infinitesimals:

$$N \operatorname{var}(\hat{m}) \sim m, \quad N \operatorname{var}(\hat{a}^{(1)}) \sim 2, \quad N \operatorname{var}(\hat{a}^{(2)}) \sim 4(e^m - 1 - m)/m^2,$$
$$N \operatorname{cov}(\hat{m}, \hat{a}^{(1)}) \sim 0, \quad N \operatorname{cov}(\hat{m}, \hat{a}^{(2)}) \sim 0, \quad N \operatorname{cov}(\hat{a}^{(1)}\hat{a}^{(2)}) \sim 2,$$
$$N \operatorname{var}(T) \sim 6m^3, \quad N \operatorname{var}(U) \sim 4\{e^m - 1 - m - \tfrac{1}{2}m^2\}.$$

2·4. *Negative binomial distribution*

$$P_r = (1+a)^{-(m/a)} \left(\frac{a}{1+a}\right)^r \Gamma\left(\frac{m}{a} + r\right) \Big/ r!\,\Gamma\left(\frac{m}{a}\right) \quad (r \geqslant 0),$$

$$\kappa_r = m \sum_{s=1}^{r} \frac{\Delta^s 0^r}{s} a^{s-1}, \quad \kappa_{r+1} = a(1+a) \frac{\partial}{\partial a} \kappa_r, \quad \kappa_{[r]} = (r-1)!\, m a^{r-1}.$$

Estimation by Method 1

$$\hat{m} = k_1, \quad \hat{a}^{(1)} = (k_2/k_1) - 1,$$

$$N \operatorname{var}(\hat{m}) = m(1+a), \quad N \operatorname{var}(\hat{a}^{(1)}) \sim 2(1+a)^2 + a(1+a)(2+3a)/m, \quad N \operatorname{cov}(\hat{m}, \hat{a}^{(1)}) \sim a(1+a).$$

Estimation by Method 2

$\hat{m} = k_1$, $\hat{a}^{(2)}$ is given implicitly as

$$\hat{a}^{(2)}/\ln(1+\hat{a}^{(2)}) = k_1/\ln(N/n_0),$$

$$N \operatorname{var}(\hat{a}^{(2)}) \sim \frac{a^2(1+a)}{m} + (1+a)^2 a^4 \left\{(1+a)^{m/a} - 1 - \frac{m}{1+a}\right\} \Big/ m^2\{a - (1+a)\ln(1+a)\}^2,$$

$$N \operatorname{cov}(\hat{m}, \hat{a}^{(2)}) \sim a(1+a).$$

Let $k_1/\ln(N/n_0) = c$, then $\hat{a}^{(2)}$ is the root of the equation

$$f(a) = a - c\ln(1+a) = 0.$$

A unique non-negative solution exists if $c \geqslant 1$. An iterative formula based on Newton's method of tangential iteration is

$$a_{i+1} = c\,\frac{(1+a_i)\ln(1+a_i)-a_i}{1+a_i-c},$$

where a_i is monotonic decreasing towards $\hat{a}^{(2)}$ if a_0 is chosen so that $f(a_0)$ is small and positive.

A one-page table of $a/\ln(1+a) = c$ makes the task of finding $\hat{a}^{(2)}$ much easier. Table 8 gives values of $\sqrt[3]{a}$ for $\log_{10} c = 0{\cdot}05(0{\cdot}01)\,0{\cdot}65$. Linear interpolation may be used for $\log_{10} c > 0{\cdot}20$, giving values of a correct to four significant figures with a possible error of one unit in the last place.

Table 8. *$\sqrt[3]{a}$ in terms of $\log_{10}c$, where a is such that $a/\ln(1+a) = c$*

$\log_{10}c$	$\sqrt[3]{a}$	Δ'	Δ''	$\log_{10}c$	$\sqrt[3]{a}$	Δ'	$\log_{10}c$	$\sqrt[3]{a}$	Δ'	$\log_{10}c$	$\sqrt[3]{a}$	Δ'
		+	−			+			+			+
0·05	0·6330	440		0·20	1·1087	257	0·35	1·4778	241	0·50	1·8445	251
0·06	0·6770	403	37	0·21	1·1344	254	0·36	1·5019	241	0·51	1·8696	252
0·07	0·7173	376	27	0·22	1·1598	252	0·37	1·5260	242	0·52	1·8948	254
0·08	0·7549	353	23	0·23	1·1850	250	0·38	1·5502	242	0·53	1·9202	254
0·09	0·7902	336	17	0·24	1·2100	248	0·39	1·5744	242	0·54	1·9456	256
0·10	0·8238	322	14	0·25	1·2348	247	0·40	1·5986	243	0·55	1·9712	257
0·11	0·8560	310	12	0·26	1·2595	245	0·41	1·6229	243	0·56	1·9969	258
0·12	0·8870	299	11	0·27	1·2840	244	0·42	1·6472	244	0·57	2·0227	260
0·13	0·9169	292	7	0·28	1·3084	244	0·43	1·6716	244	0·58	2·0487	261
0·14	0·9461	284	8	0·29	1·3328	242	0·44	1·6960	246	0·59	2·0748	262
0·15	0·9745	278	6	0·30	1·3570	242	0·45	1·7206	246	0·60	2·1010	264
0·16	1·0023	272	6	0·31	1·3812	242	0·46	1·7452	247	0·61	2·1274	266
0·17	1·0295	268	4	0·32	1·4054	242	0·47	1·7699	248	0·62	2·1540	267
0·18	1·0563	264	4	0·33	1·4296	241	0·48	1·7947	248	0·63	2·1807	269
0·19	1·0827	260	4	0·34	1·4537	241	0·49	1·8195	250	0·64	2·2076	270
0·20	1·1087			0·35	1·4778		0·50	1·8445		0·65	2·2346	

In order to interpolate for subtabular values, use Bessel's formula:

$$f_\theta = f_0 + \theta\Delta'_{\frac{1}{2}} - \frac{\theta(1-\theta)}{4}(\Delta''_0 + \Delta''_1).$$

Except where given, second-order differences may be neglected.

The exponent of the negative binomial distribution, defined by

$$k = m/a,$$

could have been used instead of the parameter a. The equation for the estimate $\hat{a}^{(2)}$ is simpler in practice than the equivalent equation for $\hat{k}^{(2)} = \hat{m}/\hat{a}^{(2)}$. It may be noted that for the negative binomial distribution both $\hat{k}^{(1)} = \hat{m}/\hat{a}^{(1)}$ and $\hat{k}^{(2)}$ are uncorrelated in large samples with the fully efficient estimator $\hat{m}$.

Use of a combined estimate

It may happen that neither Method 1 nor Method 2 is particularly efficient, and in this case the possibility may be considered of combining the two estimates $\hat{a}^{(1)}$ and $\hat{a}^{(2)}$ in order to get a more efficient estimate, $\hat{a}_w$ say, defined by

$$\hat{a}_w = \hat{a}^{(1)} + w(\hat{a}^{(2)} - \hat{a}^{(1)}).$$

The variance of $\hat{a}_w$ is a minimum for

$$w = \frac{\operatorname{var}(\hat{a}^{(1)}) - \operatorname{cov}(\hat{a}^{(1)}\hat{a}^{(2)})}{\operatorname{var}(\hat{a}^{(1)}) - 2\operatorname{cov}(\hat{a}^{(1)}\hat{a}^{(2)}) + \operatorname{var}(\hat{a}^{(2)})}.$$

The value of the minimum variance is

$$\operatorname{var}(\hat{a}_w) = \{w^2 \operatorname{var}(\hat{a}^{(2)}) - (1-w)^2 \operatorname{var}(\hat{a}^{(1)})\}/(2w-1),$$

and for the negative binomial distribution,

$$N \operatorname{cov}(\hat{a}^{(1)}, \hat{a}^{(2)}) \sim \frac{a^2(1+a)}{m}\left\{1 - \frac{m+a}{a-(1+a)\ln(1+a)}\right\}.$$

Contours of this weighting function w are given in Fig. 1. The bounding contours $\mathscr{E}_1 = 90\ \%$ and $\mathscr{E}_2 = 90\ \%$ given in this figure refer to the efficiencies of $\hat{k}^{(1)}$ and $\hat{k}^{(2)}$ as estimators of k. They are redrawn from Fig. 1 of Anscombe (1950). The contour $\mathscr{E}_w = 90\ \%$ refers to the efficiency of the estimator $\hat{k}_w = \hat{m}/\hat{a}_w$, and indicates the expected increase in efficiency.

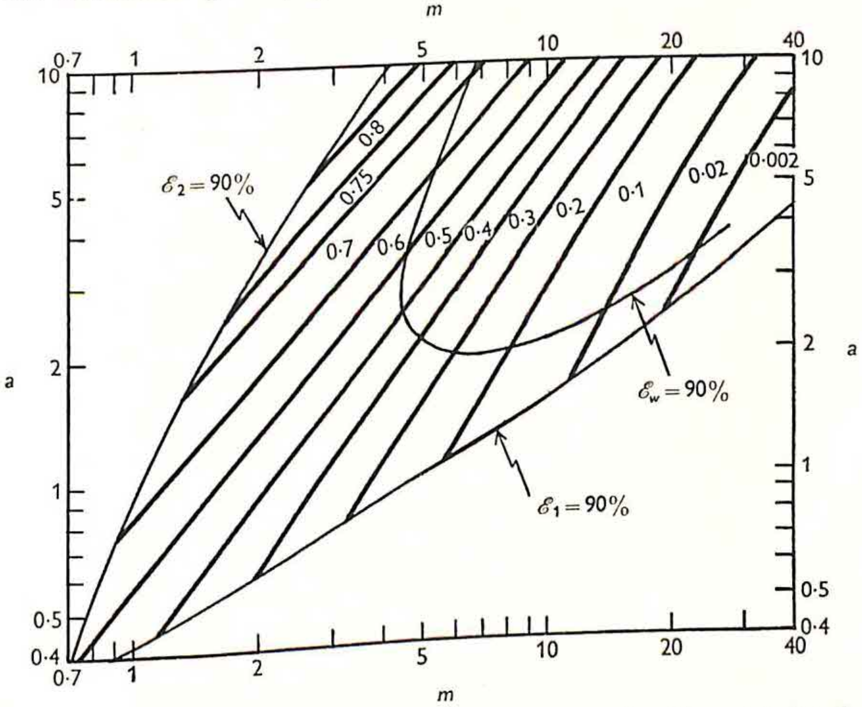


Fig. 1. Contours of the weighting function w for the negative binomial distribution.

If estimation of parameters is by Method 1, we may compute

$$T = k_3 + k_2 - 2k_2^2/k_1,$$

and test the significance of its deviation from zero by

$$N \operatorname{var}(T) \sim 2m(m+a)(1+a)^2\{2a(5a+3) + 3m(1+a)\}.$$

If estimation is by Method 2, we may compute

$$U = k_2 - k_1(1 + \hat{a}^{(2)}),$$

and test the significance of its deviation from zero by

$$N \operatorname{var}(U) \sim 2m(m+a)(1+a)\left\{\frac{(1+a)^2\ln(1+a) - a(1+2a)}{(1+a)\ln(1+a) - a}\right.$$
$$\left. + \frac{(1+a)a^4}{[(1+a)\ln(1+a) - a]^2}\{(1+a)^{1+(m/a)} - (m+1+a)\}\right\}.$$

Contours of the standard errors of T and U for $N = 100$ are given in Fig. 2. The line EE in Figs. 2, 3 and 4 represents the contour along which the efficiencies of Methods 1 and 2 for estimating the parameters of the distribution concerned are equal. Method 1 is to be preferred in the area to the right of this line, and Method 2 to the left. Logarithmic interpolation is satisfactory. For sample sizes other than $N = 100$ the standard errors stated must be multiplied by $10/\sqrt{N}$.

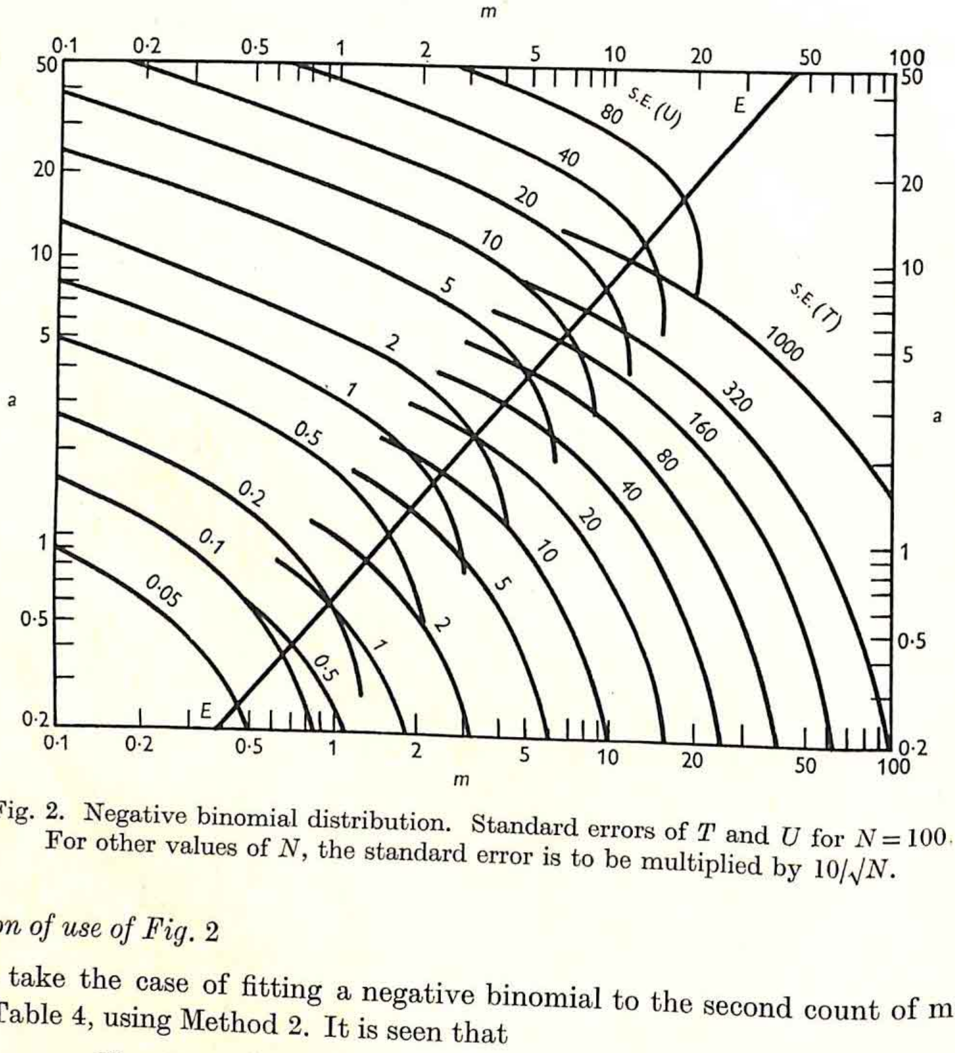


Fig. 2. Negative binomial distribution. Standard errors of T and U for $N = 100$. For other values of N, the standard error is to be multiplied by $10/\sqrt{N}$.

Illustration of use of Fig. 2

Let us take the case of fitting a negative binomial to the second count of moth eggs given in Table 4, using Method 2. It is seen that

$$N = 782, \quad \hat{m} = k_1 = 5{\cdot}75, \quad \hat{a}^{(2)} = 5{\cdot}07, \quad 10/\sqrt{N} = 0{\cdot}358.$$

The chart shows that the point (m, a) falls in the region where Method 2 is preferable and that for $N = 100$, S.E. (U) is approximately equal to 7. Consequently the estimated standard error of U for $N = 782$ is $7 \times 0{\cdot}358 = 2{\cdot}5$. The value calculated from the formula given above is 2·6.

2·5. *The Pólya-Aeppli distribution*

$$P_0 = \exp\left[-\frac{2m}{2+a}\right], \quad P_r = P_0\left[\frac{4m}{(2+a)^2}\right]^r \left\{\sum_{s=0}^{r-1} \frac{\binom{r-1}{s} X^s}{(r-s)!}\right\} \quad (r \geqslant 1),$$

where

$$X = \tfrac{1}{4}a(2+a)/m.$$

$$\kappa_r = m \sum_{s=1}^{r} a_s(r)\, a^{s-1},$$

where
$$a_s(r) = \sum_{u=1}^{s} (-)^{u+s} \binom{s}{u} u^r/2^{s-1}.$$

$$\kappa_{[r]} = r!\, m(a/2)^{r-1}, \quad \kappa_{r+1} = \left(1+a+\frac{a(2+a)}{2}\frac{\partial}{\partial a}\right)\kappa_r.$$

The probabilities P_r may conveniently be calculated from the recurrence relation

$$t_r = \left(2+\frac{Y}{r-1}\right)\alpha t_{r-1} - \alpha^2 t_{r-2} \quad (r \geqslant 2),$$

where
$$t_r = rP_r, \quad \alpha = a/(2+a) \quad \text{and} \quad Y = 4m/a(2+a).$$

Alternatively,
$$P_r = P_0 \alpha^r Y e^{-Y} M(r+1, 2, Y) \quad (r \geqslant 1),$$

where $M(\alpha, \gamma, x)$ is the confluent hypergeometric function defined by

$$M(\alpha,\gamma,x) = 1+\frac{\alpha}{\gamma}\frac{x}{1!}+\frac{\alpha(\alpha+1)}{\gamma(\gamma+1)}\frac{x^2}{2!}+\dots.$$

A table of $M(\alpha, 2, x)$ for $\alpha = 1(1)4$ and for $x = 0{\cdot}00$ by various increments up to $x = 8{\cdot}0$, was given in the *British Association Report* for 1927, p. 233; and more recently Rushton (1951) has given a table for $\alpha = 2(1)40$.

Values of the cumulative probability function can be found from the relation

$$\sum_{r=0}^{R} P_r = P_0 + P_0 \sum_{r=0}^{R} \frac{1}{r!}\left(\frac{2m}{2+a}\right)^r I_{2/(2+a)}(r, R-r+1) \quad (R \geqslant 1),$$

where $I_x(p,q)$ is the incomplete β-function.

The following approximate relations for the cumulative sums of P_r and of rP_r were used for checking purposes:

$$\frac{P_{R+1}-(2\alpha-1)P_R}{(1-\alpha)^2} < 1 - \sum_{0}^{R-1} P_r < \frac{P_{R+1}-\{(2+(Y-2)/R)\alpha-1\}P_R}{(1-\alpha)^2-\alpha(Y-2+2\alpha)/R},$$

and also
$$\sum_{0}^{R-1} rP_r = m\sum_{0}^{R-1} P_r + (R-1)P_{R-1}(\tfrac{1}{2}a)^2 - RP_R(1+\tfrac{1}{2}a)^2.$$

By writing down the log-likelihood function L and equating the quantity

$$2m(1+a)\frac{\partial L}{\partial m} + a(2+a)\frac{\partial L}{\partial a}$$

to zero, one obtains the solution $\hat{m} = k_1$. Because the probabilities P_r are in the form of a series, it does not seem possible to obtain convenient expressions for $\hat{a}$ and its large-sample variance.

Estimation by Method 1

We have
$$\hat{m} = k_1, \quad \hat{a}^{(1)} = (k_2/k_1) - 1,$$

$$N\,\text{var}\,(\hat{m}) = m(1+a), \quad N\,\text{var}\,(\hat{a}^{(1)}) \sim 2(1+a)^2 + a(1+a)(2+a)/m,$$

$$N\,\text{cov}\,(\hat{m}, \hat{a}^{(1)}) \sim \tfrac{1}{2}a(2+a).$$

Estimation by Method 2

We have
$$\hat{m} = k_1, \quad \hat{a}^{(2)} = \{2k_1/\ln(N/n_0)\} - 2,$$
$$N\,\mathrm{var}\,(\hat{a}^{(2)}) \sim \left(\frac{2+a}{2m}\right)^2 \{(2+a)^2\,(e^{2m/(2+a)} - 1) - 4m\},$$
$$N\,\mathrm{cov}\,(\hat{m}, \hat{a}^{(2)}) \sim \tfrac{1}{2}a(2+a), \quad N\,\mathrm{cov}\,(\hat{a}^{(1)}, \hat{a}^{(2)}) \sim (2+a)^2\,(m+a)/(2m).$$

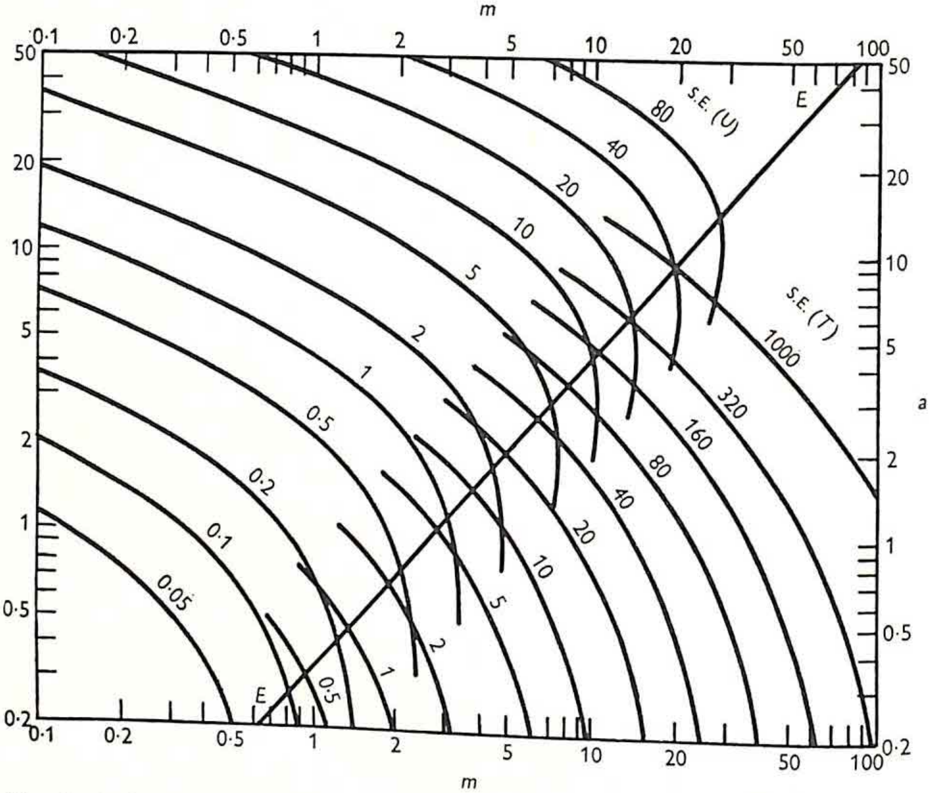


Fig. 3. Pólya-Aeppli distribution. Standard errors of T and U for $N = 100$. For other values of N, the standard error is to be multiplied by $10/\sqrt{N}$.

If estimation of parameters is by Method 1, we may compute
$$T = k_3 - k_1(3k_2^2 - k_1^2)/2k_1^2,$$
and test the significance of its deviation from zero by
$$N\,\mathrm{var}\,(T) \sim 6(1+a)^3\,m^3 + 9a(2+a)\,m\{(5a^2 + 10a + 4)\,m + a(1+a)\,(2+a)\}/4.$$
If estimation is by Method 2, we may compute
$$U = k_2 - k_1(1 + \hat{a}^{(2)}),$$
and test this by
$$N\,\mathrm{var}\,(U) \sim \frac{(2+a)^4}{4}\{e^{2m/(2+a)} - 1\} - 2(1+a)\,(2+a)\,m + (a^2 - 2)\,m^2.$$

Contours of the standard errors of T and U for $N = 100$ are given in Fig. 3. This chart can be used to give rough values of the standard errors, as illustrated above in the case of Fig. 2.

2·6. *The Neyman Type A distribution*

$$P_r = \frac{a^r}{r!}\,e^{-(m/a)} \sum_{j=0}^{\infty} \frac{j^r}{j!}\left(\frac{m\,e^{-a}}{a}\right)^j \quad (r \geqslant 0).$$

Beall (1940) suggests that for computation this expression is more conveniently put into a recurrent form. It may easily be verified that

$$P_{r+1} = \frac{m\,e^{-a}}{r+1} \sum_{u=0}^{r} \frac{a^u}{u!} P_{r-u} \quad (r \geqslant 0).$$

The following formula gives a means of checking the calculations:

$$P_r = \frac{a^{r-1}}{r!} P_1 \sum_{j=1}^{r} \frac{\Delta^j 0^r}{j!} A^{j-1}, \quad \text{where} \quad A = m\,e^{-a}/a.$$

It was also found useful to accumulate the sums of P_r and of rP_r, which are monotonically increasing towards 1 and m respectively. Approximate relations such as

$$\sum_1^R rP_r + (R+1)\left(1 - \sum_1^R P_r\right) < m$$

are easy to derive.

Alternatively, by using a table of the cumulative Poisson distribution, such as that given by Molina (1947), we have that

$$\sum_{r=0}^{c-1} P_r = 1 - e^{-(m/a)} \sum_{j=0}^{\infty} P(c, aj) \frac{(m/a)^j}{j!},$$

where

$$P(c, aj) = \sum_{i=c}^{\infty} \frac{e^{-(aj)}(aj)^i}{i!},$$

$$\kappa_r = m \sum_{s=1}^{r} \frac{\Delta^s 0^r}{s!} a^{s-1}, \quad \kappa_{r+1} = \left(1 + a + a\frac{\partial}{\partial a}\right)\kappa_r, \quad \kappa_{[r]} = ma^{r-1}.$$

By writing down the log-likelihood function L and equating the quantity

$$m(1+a)\frac{\partial L}{\partial m} + a\frac{\partial L}{\partial a}$$

to zero, we get the solution $\hat{m} = k_1$. It is not possible to obtain convenient expressions for $\hat{a}$ and its large-sample variance because of the series form of the probabilities P_r. Shenton (1949) has given an upper bound to the efficiency of the method of moments.

Estimation by Method 1

We have

$$\hat{m} = k_1, \quad \hat{a}^{(1)} = (k_2/k_1) - 1,$$

$$N \operatorname{var}(\hat{m}) = m(1+a), \quad N \operatorname{var}(\hat{a}^{(1)}) \sim 2(1+a)^2 + a(2+a)/m,$$

$$N \operatorname{cov}(\hat{m}, \hat{a}^{(1)}) \sim a.$$

Estimation by Method 2

We have $\hat{a}^{(2)}$ as the solution of the following equation in a:

$$\frac{1 - e^{-a}}{a} = \frac{\ln(N/n_0)}{k_1}.$$

The iterative formula

$$a_{i+1} = c - c\,e^{-a_i},$$

where $c = k_1/\ln(N/n_0)$, may be used. It is easily seen that a_i is monotonic increasing or decreasing towards the solution $\hat{a}^{(2)}$ if a_0 is respectively less than or greater than $\hat{a}^{(2)}$:

$$N \operatorname{var}(\hat{a}^{(2)}) \sim \frac{a^2}{m(1+a)} + \frac{a^4}{m^2\{(1+a)e^{-a} - 1\}^2}\left\{e^{m(1-e^{-a})/a} - 1 - \frac{m}{1+a}\right\},$$

$$N \operatorname{cov}(\hat{m}, \hat{a}^{(2)}) \sim a, \quad N \operatorname{cov}(\hat{a}^{(1)}, \hat{a}^{(2)}) \sim a^2\{e^{-a} - 1 - m\}/m\{(1+a)e^{-a} - 1\}.$$

If estimation of parameters is by Method 1, we may compute

$$T = k_3 - k_2 + k_1 - k_2^2/k_1,$$

and test the significance of its deviation from zero by

$$N\,\mathrm{var}\,(T) \sim 6(1+a)^3\,m^3 + a(18+50a+25a^2+2a^3) + 2a^2(3+a)\,m.$$

If estimation is by Method 2, we may compute

$$U = k_2 - k_1(1+\hat{a}^{(2)}),$$

and $$N\,\mathrm{var}\,(U) \sim 2m^2\{(1+a)^2 - (a^2/\beta)\} + ma\left\{2+a-\frac{a}{1+a}\left(\frac{(1+a)\,\alpha}{(1+a)\,\alpha-1}\right)^2\right\} + \frac{a^4}{\beta^2}\{e^{\alpha m}-1\},$$

where $\alpha = (1-e^{-a})/a$ and $\beta = 1-(1+a)\,e^{-a}$. Contours of the standard errors of T and U for $N = 100$ are given in Fig. 4, which may be used as illustrated in the case of Fig. 2.

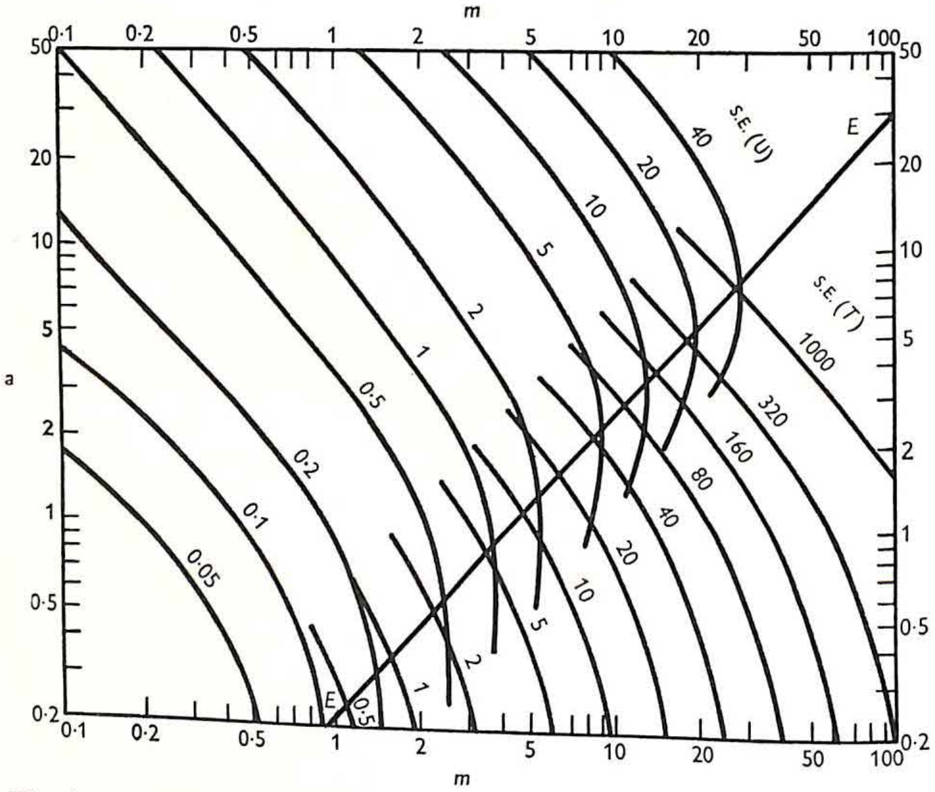


Fig. 4. Neyman Type A distribution. Standard errors of T and U for $N = 100$. For other values of N, the standard error is to be multiplied by $10/\sqrt{N}$.

This subject of research was suggested by Mr F. J. Anscombe, whom I have much pleasure in thanking for his helpful suggestions and criticisms. I would like to thank Mr D. A. East for preparing the figures for publication, and for the computing involved in tabulating the raw data given by Beall (1938) and Marshall (1936). I also wish to thank Mrs E. H. Laurie for help in computing the contours of S.E. ($U_{\mathrm{P.}}$). I am grateful for the generous way in which Dr E. E. A. Archibald and Dr H. Barnes have made available, in private correspondence, fuller details of their published data, and for their permission to publish the material.

The first draft of this work was completed during the tenure of a post-graduate research grant made by the Department of Scientific and Industrial Research. The investigations leading to the results given in Tables 7A and 7B were made possible by a grant from King's College Research Fund.

REFERENCES

ANSCOMBE, F. J. (1950). *Biometrika*, **37**, 358.
ARCHIBALD, E. E. A. (1948). *Ann. Bot., Lond.*, **12**, 221.
BARNES, H. & STANBURY, F. A. (1951). *J. Ecol.* **39**, 171.
BEALL, G. (1938). *Biometrika*, **30**, 422.
BEALL, G. (1940). *Ecology*, **21**, 460.
CLAPHAM, A. R. (1936). *J. Ecol.* **24**, 232.
HANSON, H. C. (1934). *J. Agric. Res.* **49**, 815.
MARSHALL, J. (1936). *Ann. Appl. Biol.* **23**, 150.
MOLINA, C. E. (1947). *Poisson's Exponential Binomial Limit.* N.Y.: Van Nostrand.
RUSHTON, S. (1951). Certain sequential tests of composite hypotheses. London University, Ph.D. Thesis.
SHENTON, L. R. (1949). *Biometrika*, **36**, 450.
STEIGER, T. L. (1930). *Ecology*, **11**, 170.
STIRRET, G. M., BEALL, G. & TIMONIN, M. (1937). *Sci. Agric.* **17**, 587.

MISCELLANEA

Time intervals between accidents—a note on Maguire, Pearson and Wynn's paper

By G. A. BARNARD, *Imperial College, London*

Our object in this note is to point out the possibility of using Birnbaum's recent (1952) tabulation of the distribution of Kolmogoroff's D-statistic in connexion with some of the problems discussed by Maguire, Pearson & Wynn (1952) in their paper on industrial accidents. In Table 1 of the latter paper we are given the time intervals between accidents involving more than ten men killed for the period 6 December 1875 to 29 May 1951. The question arises whether the data are consistent with the assumption that the accident rate over the period has been uniform.

Let us begin by labelling $i = 1, 2, \ldots, n$ the n ($= 109$) accidents considered, in a random manner (i.e. not, for example, according to their order of occurrence). Since the accident labelled i is equally likely to be any one of the accidents, it follows that if the accident rate is uniform, this accident is equally likely to occur anywhere in the time interval considered. In other words, if x_i denotes the time, measured from the start of the interval, at which the accident i occurred, x_i is an observation from a rectangular distribution with range from 0 to T, the total length of the interval. And if the accidents are independent of one another, the x_i are n independent observations from this rectangular distribution. If we form from these n observations the sample cumulative distribution $F_n(t)$ in the manner indicated by Birnbaum (1952), we have

$$nF_n(t) = \text{no. of accidents occurring at or before time } t,$$

while for the theoretical cumulative we have

$$nF(t) = n\,.\,t/T,$$

and so if D_n denotes Kolmogoroff's statistic,

$$\begin{aligned} nD_n &= n \max_{0<t<T} |\,F_n(t) - F(t)\,| \\ &= \max_{0<t<T} |\,(\text{no. of accidents up to time } t) - nt/T\,|. \end{aligned}$$

To estimate nD_n from Maguire *et al.* Table 1, we notice that the time intervals are, with very few exceptions, below expectation until the 38th accident, which occurred 4191 days from the beginning. At this point we have

$$\begin{aligned} 109\,.\,|\,F_{109}(4191) - F(4191)\,| &= 38 - (109 \times 4191/26263) \\ &= 20{\cdot}6, \quad \text{approximately.} \end{aligned}$$

Hence nD_n is at least 20·6 for our data. Now according to Birnbaum the approximation $1{\cdot}6276\sqrt{n}$ is good for the 1 % point of nD_n provided n is larger than 80. We find $1{\cdot}63\sqrt{109} = 17{\cdot}0$, approximately, so that the 1 % point is exceeded, and the data show a significant departure from the hypothesis tested. A lengthier calculation shows that nD_n is about 25, in fact. Either the accidents are not independent of each other, or the accident rate is not uniform.

We may now wish to examine the hypothesis that the expected number of accidents in the interval $(t, t+dt)$ instead of being $\lambda\,dt$, where λ is a constant, is $\lambda(t)\,dt$, $\lambda(t)$ being a function of t. We retain the assumption that the accidents are independent of each other. If we introduce a new 'time' variable u by the condition $du = \lambda(t)\,dt$, it is evident that the accident rate will again be uniform referred to u-time. Thus the argument given above applies, with the result that we now have

$$nD_n = \max_{0<t<T} \left|(\text{no. of accidents up to time } t) - n\int_0^t \lambda(x)\,dx \Big/ \int_0^T \lambda(x)\,dx\right|.$$

If, for example, we try $\lambda(t)$ proportional to e^{-kt}, the second term within the modulus sign becomes

$$109\,.\,(1-e^{-kt})/(1-e^{-26263k}),$$

and we find that nD_n is well within the 95 % limit (given by Birnbaum as $1{\cdot}3581\sqrt{n}$) if we take $k = 1/20000$. This corresponds to a logarithmic decrement of about 1·8 % per annum in the accident rate. Thus the data are consistent with the possibility of an exponential decrease in the accident rate. Corresponding to any given probability level there will evidently be a highest and a lowest value of k which will be consistent with the data, at this probability level. These two values of k will give a confidence interval for k, assuming the decrease to be exponential in form. Similar arguments will apply to any other parametric family of decremental curves.

REFERENCES

Birnbaum, Z. W. (1952). Numerical tabulation of Kolmogoroff's statistic for finite sample size. *J. Amer. Statist. Ass.* **47**, 425–41.

Maguire, B. A., Pearson, E. S. & Wynn, A. H. A. (1952). The time intervals between industrial accidents. *Biometrika*, **39**, 168–80.

Further notes on the analysis of accident data

By B. A. MAGUIRE, E. S. PEARSON and A. H. A. WYNN

1. We are very glad that Prof. Barnard (1953) has drawn attention in the preceding paper to the fact that the Kolmogoroff test may be used very simply to establish departure from randomness in a series of events occurring in sequence, either in time or space*. The figure reproduced below illustrates diagrammatically the application of the test to the accident data from our earlier paper (Maguire, Pearson & Wynn, 1952) used by Barnard. If t_i, measured in days from 6 December 1875, represents the time of occurrence of the ith accident following that which occurred at $t = 0$, then the 110 cumulative sample points (t_i, i), $i = 0, 1, \ldots, 109$, have been plotted (section A–B of the chart). A central, continuous line joins the first and last points $(0, 0)$ and (t_n, n), where $n = 109$, $t_{109} = T = 26{,}263$; this is the theoretical cumulative line. Two parallel 'control' lines have been drawn on each side of this line at distances (measured parallel to the axis of i) of $\pm 1{\cdot}6276\sqrt{n} = \pm 17{\cdot}0$ forming a significance belt.

2. The Kolmogoroff theorem states that if the accidents have occurred at random with constant expectation during the period T, then the probability is approximately only 0·01 that the track of the cumulative points (t_i, i) will pass outside this belt. The points cross the upper limit of the belt where $i = 34$, $t = 3931$. Birnbaum (1952) has shown that when $n \geqslant 100$ there is good agreement as far out in the tail as the 1 % point between the true and the limiting distributions of the Kolmogoroff statistic. If we may assume that the approximation is also adequate at more extreme limits, it is possible to make use of Smirnoff's (1939, 1948) table of the limiting probability integral. This shows that

$$nD_n' = \text{maximum}\,|\,nt_i/T - i\,|,$$

which is 24·6 at $i = 53$, is also significant at the 0·001 % level. It is quite clear, therefore, as Barnard points out, that there is evidence of very significant departure from randomness, the average interval having increased during the period.

3. Before following up one or two points concerning methods of analysis suggested by Barnard's note, it is perhaps desirable to say a few words about the particular accident data of this example. The figures are taken from Table 1 of our 1952 paper and show the intervals between mining accidents in Great Britain due to explosions involving more than ten men killed, during the 75-year period 1875–1950. They were introduced to illustrate certain points in our discussion—for example, the use of the following two tests of the hypothesis that the overall distribution of 109 intervals was consistent with sampling from a single exponential population:

(*a*) the M-test (heterogeneity of variance) applied to single intervals,

(*b*) Fisher's g-test, based on the ratio of the longest interval to the average interval.

Neither test established a significant departure from the exponential law.

4. Although we might and perhaps should have proceeded further, we did not in fact attempt to analyse these data further in the paper, i.e. we did not examine for changes in accident expectation within the period. The result brought out by the Kolmogoroff test, namely, that the accident intervals were shorter at the beginning than at the end of the period, can also be established quite simply by using another test which we gave (Maguire *et al.* 1952, p. 172) but did not use on these data. If T_1 is the time covered by the first 54 accidents and T_2 the time covered by the next 54, then

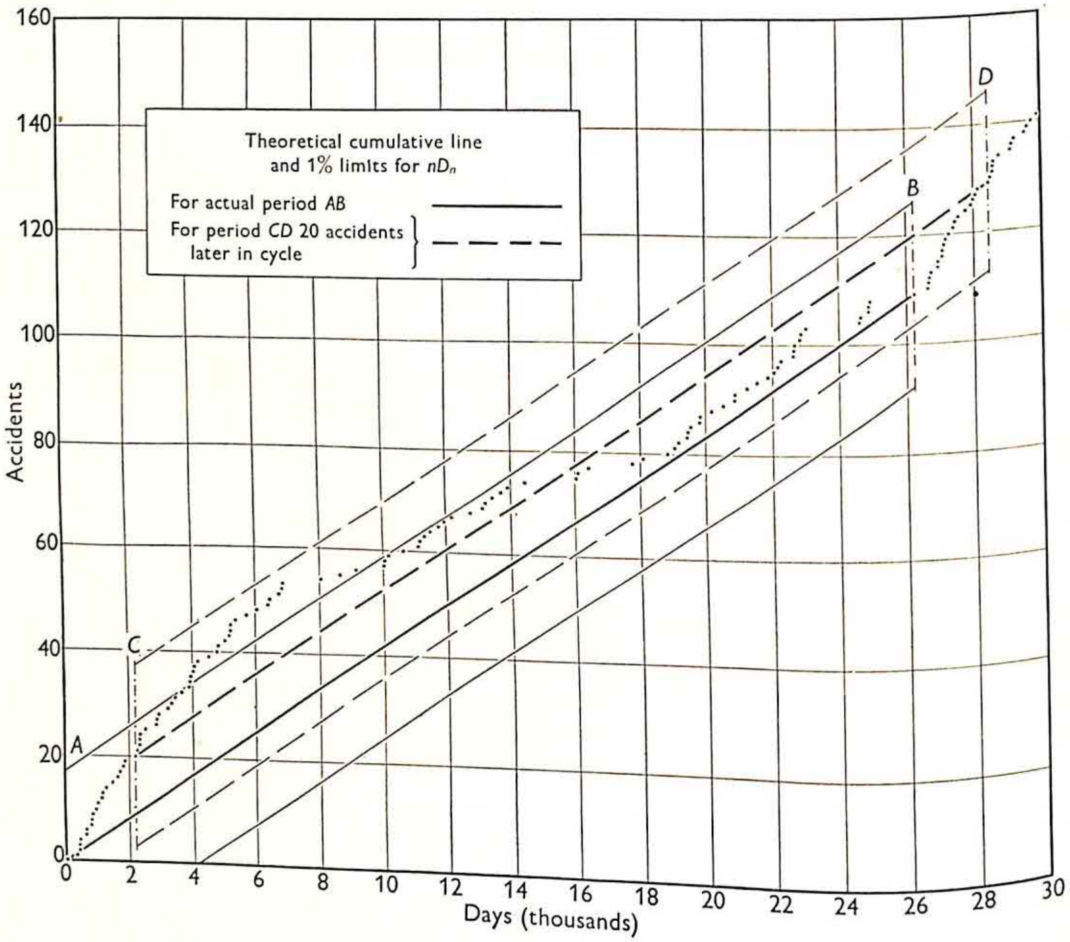
$$T_1 = t_{54} - t_0 = 8042, \quad T_2 = t_{108} - t_{54} = 16864.$$

---

* Some remarks of Bartlett (1949, p. 216) are also of interest in this connexion.

If the accidents have occurred randomly with constant expectation, T_2/T_1 should be distributed as a variance ratio, F, with degrees of freedom $\nu_1 = \nu_2 = 108$. But $F = 16864/8042 = 2{\cdot}10$, which is significant at the 0·1 % level.

5. To study the changes that may have occurred in the risk of explosions in mines would involve the consideration of a very complex story, and it would be out of place to make the attempt here. The following summary, however, gives a broad picture of what seems to have occurred. There appears to have been a substantial improvement in mine safety from 1870–90, and the average gravity of explosions, as judged from the number of casualties per explosion, was less in the period 1920–50 than between 1890–1920, but the frequency of explosions was not reduced; on the other hand, there may have been some increase in the risk of explosion in recent years in relation to coal output, the number of men in employment and the number of mines being worked, but this has not been fully investigated by the authors. There has been a considerable reduction in mining accidents in some other important categories.



6. As Barnard has pointed out, the Kolmogoroff test is likely to be efficient in detecting a monotonic change in the risk $\lambda(t)$ whether this is of exponential character or follows a linear trend. Under these conditions, the plot of the cumulative points (t_i, i) moves away from the cumulative line and then comes back again in a single broad sweep. If $\lambda(t)$ does not change monotonically the position may, however, be very different, since the track of the cumulative points may now cross the theoretical line several times and never reach the boundary of the belt. For example, under certain conditions $\lambda(t)$ might fluctuate as an autoregressive function; we might then arrange the sequence of intervals in a closed cycle and look for a test for detecting changes in $\lambda(t)$ which is independent of any particular starting-point in the cycle.

7. In the diagram we have introduced this idea by adding the 1st, 2nd, 3rd, ..., etc., intervals after the 109th. If now we apply the Kolmogoroff test to the stretch CD instead of AB, i.e. to the 110 accidents of the cycle starting at the 20th accident of the original series, the expectation line and the two parallel 0·5 % limits are as shown by broken lines. The track of cumulative sample points now never passes

outside the 99 % belt and significance would not be established by the test. The maximum value of $|nt_i/T - i|$ is now 13·3, which (using the limiting distribution of Smirnoff) falls near the 8 % significance level.

8. Although the calculation is somewhat laborious with n so large, it is of interest to consider the application of a further test which is allied to that of Kolmogoroff, namely, the ω^2-test developed by Cramér (1928), von Mises (1931) and Smirnoff (1936). If $F(x)$ is the continuous cumulative frequency function of a random variable x specified by the hypothesis tested, and if $x_1, x_2, \ldots, x_n$ are n observed values of x arranged in ascending order of magnitude, then

$$\omega^2 = \frac{1}{12n^2} + \frac{1}{n} \sum_{i=1}^{n} \left\{ F(x_i) - \frac{2i-1}{2n} \right\}^2. \tag{1}$$

For the case where the distribution of $x = t$ is rectangular in the interval $(0, T)$, $F(x_i) = F(t_i) = t_i/T$ and we may write

$$n\omega^2 = \frac{1}{12n} + \frac{1}{n^2} \sum_{i=1}^{n} \{nt_i/T - (i - 0{\cdot}5)\}^2. \tag{2}$$

Thus, while the Kolmogoroff test uses the maximum value of n times the difference $t_i/T - i/n$, the ω^2 test uses the sum of squares of the n differences $d_i = t_i/T - (2i-1)/(2n)$. It will be seen that if the pattern of the n events is reproduced on a line of unit length, then d_i is the amount by which the ith point is displaced from the corresponding point in the regular series

$$1/(2n),\ 3/(2n),\ \ldots\ (2i-1)/(2n),\ \ldots\ (2n-1)/(2n).$$

Thus if denser concentrations of points alternate with stretches of rarer occurrence, ω^2 may increase significantly above expectation.

9. The first two moments of $n\omega^2$ have been known for some time, but we are indebted to Mr B. A. M. Thomas* for expressions for the 3rd and 4th moments about the mean. Thus we have for the first four moments:

$$\left.\begin{aligned} \mu_1' &= \frac{1}{6}, & \mu_2 &= \frac{4n-3}{180n}, \\ \mu_3 &= \frac{32n^2 - 61n + 30}{3780n^2}, & \mu_4 &= \frac{496n^3 - 1532n^2 + 1671n - 630}{75600n^3}. \end{aligned}\right\} \tag{3}$$

The distribution is very far from normal. It will be found that when $n \to \infty$ the limiting values of the moments agree with the values of the cumulants of the limiting distribution of $n\omega^2$ given by Anderson & Darling (1952). The use of these moments in deriving an approximation to the distribution of $n\omega^2$ in relatively small samples requires much fuller consideration, but the following values of the standard deviation and the moment ratios $\beta_1 = \mu_3^2/\mu_2^3$ and $\beta_2 = \mu_4/\mu_2^2$, which we owe to Thomas, show that when $n = 109$ the distribution of $n\omega^2$ must be approaching the limiting form:

n	S.D.	β_1	β_2
10	0·143	5·53	11·23
20	0·146	6·03	12·24
50	0·148	6·33	12·87
100	0·149	6·43	13·08
1000	0·149	6·52	13·26
∞	0·149	6·53	13·29

We shall therefore use below the percentage points of the limiting distribution tabled by Anderson & Darling (1952, p. 203), in particular

5 %	4 %	1 %	0·1 %
0·461	0·499	0·743	1·168

---

* The results were included in an essay presented as part of the Examination for the B.Sc. Special Degree of the University of London (1952).

10. We have now calculated $n\omega^2$ from equation (2)* for the $n = 109$ accident times t_i considered previously, (*a*) starting from the zero point at 6 December 1875 and (*b*) from the date of the 20th accident in the artificial cycle described in para. 7. We then find for (*a*) and (*b*):

(*a*) $n\omega^2 = 1\cdot541$, which is a value far beyond the 0·1 % significance level given above;

(*b*) $n\omega^2 = 0\cdot474$, a result significant at the 5 % (but not at the 4 %) level.

It follows that for the present data, in its original form, both the Kolmogoroff and ω^2-test give very clear evidence of departure from randomness. When, however, the intervals are arranged in cyclical form and the analysis starts at the 20th accident, the former test does not, and the latter does, establish significance at the 5 % level.

11. It would not be legitimate to make any general comparison of the two tests on the basis of these special results. One broad conclusion, however, may, we think, be drawn. As one of us has emphasized before (Pearson, 1942), while it seems attractive to transform a statistical problem into one of testing whether a sample has been drawn from a rectangular population, it is still not possible to determine the most efficient test to use for the purpose unless we can formulate the type of departure from the rectangle that is likely to arise. It appears that the Kolmogoroff test will only be powerful in detecting certain kinds of variation in $\lambda(t)$ and a similar position will hold for the ω^2-test.

REFERENCES

Anderson, T. W. & Darling, D. A. (1952). *Ann. Math. Statist.* **23**, 193–212.
Barnard, G. A. (1953). *Biometrika*, **40**, 212.
Bartlett, M. S. (1949). *J. Roy. Statist. Soc.* B, **11**, 211.
Birnbaum, Z. W. (1952). *J. Amer. Statist. Ass.* **47**, 425–41.
Cramér, H. (1928). *Skand. Aktuar. Tidskr.* **11**, 13–74, 141–80.
Maguire, B. A., Pearson, E. S. & Wynn, A. H. A. (1952). *Biometrika*, **39**, 168–80.
von Mises, R. (1931). *Wahrscheinlichkeitsrechnung.* Vienna: Deuticke.
Pearson, E. S. (1942). *Biometrika*, **32**, 311–16.
Smirnoff, N. V. (1936). *C.R. Acad. Sci., Paris*, **202**, 449.
Smirnoff, N. V. (1939). *Bull. Math. Univ. Moscou*, **2**, fasc. 2.
Smirnoff, N. V. (1948). *Ann. Math. Statist.* **19**, 279–81.

On a method of estimating biological populations in the field

By C. C. Craig, *University of Michigan*

If it be assumed that the probability that a member of a biological population of a given class be found in any subregion of a well-defined region is proportional to the size of the subregion, the obvious way to estimate the population of the entire region is to count the number found in any randomly chosen subregion. If the total population of the region is M, if the ratio of the size of the subregion chosen to that of the whole region is p, and if n is the number found in the subregion, then the likelihood of the sample result is

$$\binom{M}{n} p^n q^{M-n}, \quad q = 1-p$$

which is maximized for the integer (or integers) $\hat{M}$ satisfying the inequality

$$\frac{n}{p} - 1 \leqq \hat{M} \leqq \frac{n}{p}. \tag{1}$$

In large samples the variance of this result is that of n/p which is

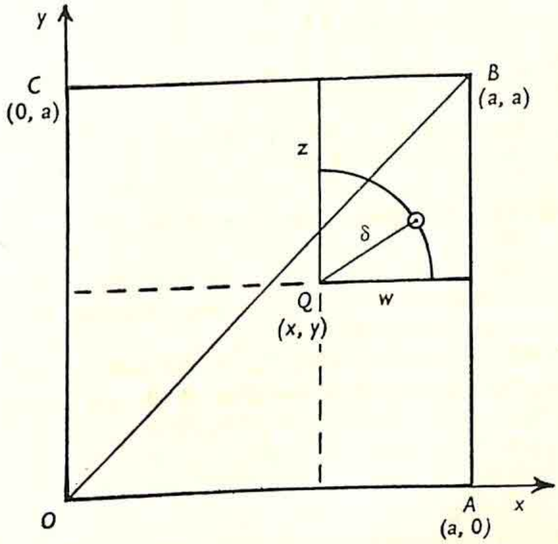
$$M\frac{q}{p}. \tag{2}$$

---

* The sampling distributions of both nD_n and ω^2 are derived on the assumption that the n values of t fall within a fixed interval $(0, T)$, i.e. they are conditional distributions for fixed T and n. We have set $T = t_n$, the period having been closed at the date of the last accident, the Easington disaster of 29 May 1951. Thus neither Barnard nor the present authors have quite kept to the rules in making their calculations, but the error is unimportant.

However, one of my biologist colleagues has proposed that a method of estimation be devised which is based on observed distances between individuals and their nearest neighbours. More definitely, let the region be a plane area and let the co-ordinates (x, y) of the position of any individual be independent stochastic variables each obeying a rectangular distribution law. An individual being chosen at random, observe the distances to his nearest neighbours in each of the four quadrants determined by parallels to the co-ordinate axes drawn through his position. From a sample distribution of such distances, how should the total population be estimated and how good an estimate can be obtained?

Since individuals are assumed to obey a constant distribution function over the area, I will choose points $Q(x, y)$ at random from which to observe distances δ to the nearest members of the population. For simplicity, I will take the region to be a square with its vertices at $O(0, 0)$, $A(a, 0)$, $B(a, a)$, $C(0, a)$, and I will consider only distances to individuals lying in the first quadrant with Q as origin. This avoids the complication that for a choice of Q the distances to the nearest neighbour in each of the four quadrants about Q will not be independent. Further, I assume that the population observed is alive and freely moving, or at any rate is such that each new choice of Q gives an independently observed value of δ.



To find the probability density function (p.d.f.) for δ there are four cases to be considered. Setting $a-x=w$ and $a-y=z$, the four cases are: (1) $\delta \leqq$ the smaller of w and z; (2) $\delta >$ the smaller of w and z but $\leqq$ the larger; (3) $\delta >$ the larger of w and z but $\leqq \sqrt{(w^2+z^2)}$; (4) $\delta > \sqrt{(w^2+z^2)}$.

For case (1) suppose first that Q lies in OAB. Then for any given Q for which now $\delta \leqq w$, if the total population in the whole region is M, the p.d.f. for δ is

$$\frac{\pi M}{2a^2}\,\delta\left(1-\frac{\pi\delta^2}{4a^2}\right)^{M-1} d\delta \quad (\delta \leqq w). \tag{3}$$

Since this is independent of w, the p.d.f. for all Q's is

$$\frac{\pi M\delta}{2a^2}\left(1-\frac{\pi\delta^2}{4a^2}\right)^{M-1}\left(1-\frac{\delta}{a}\right)^2 d\delta \quad (\delta \leqq w). \tag{4}$$

On summing this over all δ's, from 0 to a, the total frequency for case (1) is

$$1-\frac{2\Gamma(M+1)}{\Gamma(M+\frac{3}{2})}\,I_{\frac{1}{4}\pi}(\tfrac{1}{2}, M+1)+\frac{4}{M+1}\,I_{\frac{1}{4}\pi}(1, M+1)=P_1(M),$$

in which $I_x(p, q)$ is the Incomplete Beta-function of K. Pearson.

For two of the three remaining cases, namely, those in which for a chosen Q in OAB, $w<\delta\leqq z$ and $z<\delta\leqq \sqrt{(w^2+z^2)}$, it is not difficult to find the p.d.f. for δ for a fixed Q, but I have been able to express the sums over all random positions of Q only as quadratures. For the fourth case, in which no member of the population in the region is found in the first quadrant with Q as origin, the probability of such an event is

$$\int_0^a\int_0^a\left(1-\frac{wz}{a^2}\right)^M\frac{dw\,dz}{a^2}=\frac{1}{M+1}\sum_{i=1}^{M+1}\frac{1}{i}=P_4(M). \tag{5}$$

Now a sample of δ's obtained for N random choices of Q will contain k_1 δ's less than or equal to the distance from Q to the nearer of the right and upper boundaries, i.e. belonging to case (1), k_2 δ's belonging to case (2), k_3 δ's belonging to case (3), and k_4 instances in which no δ was observable. To attempt to use all the information in the sample requires that a rather complicated likelihood function be maximized. However, the short table of values of $P_1(M)$ given shows that, if M is more than 50, the use of only the portion of the sample in which the δ's are not greater than the nearer of the right and the upper boundaries utilizes on the average some three-quarters or more of the total information available.

M	$P_1(M)$	M	$P_1(M)$
10	0·5058	50	0·7442
20	0·6216	100	0·8134
30	0·6804	200	0·8652

The likelihood function based on the k_1 observations in which δ is not greater than the smaller of w and z for each Q chosen is, from (4),

$$P = \prod_{i=1}^{k_1} \frac{\pi M \delta_i}{2a^2}\left(1-\frac{\pi\delta_i^2}{4a^2}\right)^{M-1}\left(1-\frac{\delta_i}{a}\right)^2.$$

Maximizing $\log P$ with respect to M, I get as the maximum-likelihood estimate of M,

$$\hat{M} = -k_1\Big/\sum_{i=1}^{k_1}\log\left(1-\frac{\pi\delta_i^2}{4a^2}\right). \qquad (6)$$

The asymptotic variance of $\hat{M}$ is

$$\sigma_{\hat{M}}^2 = \frac{M^2}{k_1}.$$

A rapid method for estimating the correlation coefficient from the range of the deviations about the reduced major axis

By C. H. LEIGH-DUGMORE, *Dunlop Research Centre, Birmingham*

In many studies of organic correlation (in palaeontology, for example) it is useful to fit the allometric growth line of Huxley (1932) by the reduced major axis method of Kermack & Haldane (1950). This method has three advantages: first, it is more appropriate than conventional regression analysis because the biological variability of the material is usually much greater than that due to errors of measurement and because the terms 'dependent variate' and 'independent variate' have no real meaning; secondly, the reduced major axis is invariant under both rotation of axes and change of scale—the latter is more important; thirdly, its computation involves only the sums and sums of squares of the two variates and not the sum of products. When large samples are being studied (as by Parkinson, 1952) of individuals each with several measured characters such a reduction in computation is a great advantage. Kermack & Haldane give formulae for the sampling variances of the estimates of the allometric growth constants obtained by their method. Each formula involves the correlation coefficient so that, if tests of significance of the growth constants are required and the product-moment correlation has to be calculated this advantage in computation is lost.

When the distribution of the variates, or, in studies of allometry, of their logarithms, is homoscedastic and nearly normal (this is often the case), an estimate of the correlation coefficient can quite simply be made from a consideration of the range of deviations among the scatter of points about the reduced major axis.

Following, as closely as possible, the notation of Kermack & Haldane (1950) let x, y be the two variates under consideration and suppose their logarithms, X, Y, to base h to be normally distributed (in practice $h = 10$, usually; decimal are more convenient than natural logarithms). The equation to the reduced major axis of the X, Y correlation surface (the Kermack & Haldane line) is

$$\frac{Y-\bar{Y}}{X-\bar{X}} = \frac{s_Y}{s_X}. \qquad (1)$$

We will rewrite this in the form

$$\eta = \bar{Y} - \frac{s_Y}{s_X}\bar{X} + \frac{s_Y}{s_X}X, \qquad (2)$$

in which η is the estimate of Y from (1) for any given value of X. This may seem artificial, since the Kermack & Haldane line is not intended for estimation in the sense that the regression line is used; η is introduced for convenience in the following argument.

In (2)

$$\bar{Y} - \frac{s_Y}{s_X}\bar{X} = \log b, \tag{3}$$

$$\frac{s_Y}{s_X} = a, \tag{4}$$

where a, b are estimates of the parameters α, β in the allometric growth equation

$$y = \beta x^{\alpha}. \tag{5}$$

The sum of squares of the deviations of the Y-values from the line (2) is

$$\begin{aligned} S(Y-\eta)^2 &= S(Y-\bar{Y}-a(X-\bar{X}))^2 \\ &= 2(1-r)\,S(Y-\bar{Y})^2, \end{aligned} \tag{6}$$

where r is the correlation coefficient of the sample of X, Y. Hence

$$r = 1 - \frac{S(Y-\eta)^2}{2S(Y-\bar{Y})^2} \tag{7}$$

provides an estimate of the correlation, given the sums of squares of the variate Y about the Kermack & Haldane line and about the mean value, $\bar{Y}$.

In practice, $S(Y-\bar{Y})^2$ will already have been calculated and used to find a. It remains to estimate $S(Y-\eta)^2$.

When the distribution is reasonably homoscedastic and normal, an estimate of this sum of squares, sufficiently reliable for many purposes, can be obtained from the range of the sample of deviations about the fitted line (2) using a table of the mean range in different-sized samples such as Table XXII in *Tables for Statisticians and Biometricians*, Part II.

To apply this method the data are plotted, usually on logarithmic graph paper, the extreme deviations are found by inspection and an estimate of the correlation coefficient is easily calculated as in the following example. When a desk-calculating machine is not available the range method is undoubtedly more speedy than calculation of the product-moment correlation. For machine calculation it is necessary to look up the logarithms of the data; for the range method the data have to be plotted. If a plot is necessary for other purposes the range method has the advantage in speed; if a plot is otherwise unnecessary it may still have the advantage depending on the sample size and the fineness of any grouping of the data.

As an example the method can be applied to the logarithmic data of the *Micraster cor-anguinum* population, used as an illustrative example by Kermack & Haldane (1950).

If the logarithms of these data are plotted (to base 10) and the line (2) fitted (calculated from the logarithmic data) the greatest positive deviation from the line is 0·081 and the greatest negative deviation −0·125. Hence the range of the deviations is 0·206 and, since the mean range of samples of 338 from a normal population is about 5·83 times the standard deviation, we have 0·0354 as an estimate of the standard deviation of the deviations about the line. Substituting this and Kermack & Haldane's value for s_H^2 ($= s_Y^2$ when corrected to logarithms to base 10) in (7) we have a value of $r = 0{\cdot}8839$ to compare with $0{\cdot}8865 \pm 0{\cdot}0116$ given by Kermack & Haldane.

The method has been used in other examples by Dr D. Parkinson and the author and has been compared with the product-moment correlation. The agreement has been found satisfactory and the differences have not affected the results of tests of significance.

REFERENCES

HUXLEY, J. S. (1932). *Problems of Relative Growth*. London: Methuen.

KERMACK, K. A. & HALDANE, J. B. S. (1950). Organic correlation and allometry. *Biometrika*, **37**, 30.

PARKINSON, D. (1952). Ontogeny and phylogeny of the Lower Carboniferous Brachiopod *Schizophoria resupinata*. *Abstr. Proc. Geol. Soc.* no. 1484, p. 52.

The effect of overlapping in bacterial counts of incubated colonies

By C. MACK, *Shirley Institute, Manchester*

INTRODUCTION

1. To estimate the number of bacteria present in a given atmosphere a sample of the air is drawn over a plate of suitable material in such a manner that the bacteria are deposited thereon. The plate is then incubated and the bacteria multiply and form colonies. However, colonies which are close together may coalesce or one type of bacteria may inhibit the growth of another type by antibiotic effect so that the count of colonies is an underestimate of the original number of bacteria present, and this paper is concerned with the correction to be applied in such circumstances. Armitage (1949), in a paper on overlapping in dust particle counting, mentions the similarity of the bacterial colony overlap problem but does not deal with it in detail. However, there are some differences in the two problems, e.g. the colonies may cover a substantial fraction of the total area of the plate, the dust particles usually do not; furthermore, the dust particles are of small but finite size often of different shapes, while bacteria are, before incubation, good approximations to mathematical points.

There is a further difference between this paper and Armitage's in the approach adopted. Thus Armitage estimates by the first few terms of an expansion the number of 'clumps' which would be formed, on the average, by particles placed at random over the plate, the convergence of the expansion depending on the ratio of dust particle area to the total area. Though I have given formulae for the occurrence of 'aggregates' formed by points placed at random (Mack, 1948, 1949) this direct approach was rejected in favour of treating the problem as one of inverse probability. By this means straightforward exact formulae for the 'most probable' number of bacteria and the posterior probability of the 'probable' number of bacteria are obtained appropriate to each particular count. Lidwell (1948) has given a formula for the most probable number, but the analysis is based on admittedly over-simplified assumptions, and his result is too low.

The results of this paper may also be applied to dust particle counting. Being exact, they avoid difficulties associated with the use of the first few terms of a series and are applicable to counts covering any range of different particle sizes.

DERIVATION OF THE FORMULAE

2. Let q be the number of observed colonies and suppose $a_1, a_2, \ldots, a_q$ the area of each of these, A being the total area of the plate upon which bacteria could fall. Now suppose m bacteria fell originally and, for simplicity, assume they are all of the same type. Of these m bacteria one must have fallen in a_1, one in a_2, and so on up to a_q, while the remaining $m-q$ must have fallen somewhere in the areas $a_1, a_2, \ldots, a_q$.* Assuming each particle fell at random the probability $F(m)$ of such an occurrence is given by

$$F(m) = \frac{ma_1}{A}\frac{(m-1)\,a_2}{A}\frac{(m-2)\,a_3}{A}\cdots\frac{(m-q+1)a_q}{A}\left(\frac{a_1+a_2+\ldots+a_q}{A}\right)^{m-q}. \tag{1}$$

If we use Bayes's postulate that all values of m are equally likely, then, writing $r = (a_1+a_2+\ldots+a_q)/A$, the most probable number n of bacteria originally deposited is the smallest positive value of m which satisfies

$$\frac{F(m+1)}{F(m)} \equiv \frac{(m+1)\,r}{m-q+1} < 1. \tag{2}$$

Hence

$$n = [q/(1-r)], \tag{3}$$

the square bracket denoting the 'integral part of'.† Dividing $F(m)$ by a suitable factor so that the sum from $m = q$ to ∞ is unity, we find that the posterior probability $f(m)$ of m being the original number of bacteria is given by

$$f(m) = \frac{m(m-1)\ldots(m-q+1)\,r^{m-q}(1-r)^{q+1}}{q!}. \tag{4}$$

---

* The areas in which bacteria could have fallen may be somewhat less than the areas after incubation—the smaller values should thus be taken for $a_1, \ldots, a_q$.

† Formula (3) for n may also be derived by using the principle of maximum likelihood.

A similar argument to the above shows that if the prior probability were proportional to (*a*) $1/m$, (*b*) $(m+1)$, then the most probable number would be, in case (*a*) $[(q-r)/(1-r)]$, and in (*b*) $[(q+r)/(1-r)]$. Since q is usually of the order of 100 the differences are negligible in practice, so that the reasonableness of the inverse probability treatment is justified.

3. If there are several different types of bacteria, then each type may be treated separately. Thus, if there are q_1 colonies of type 1, formulae (3) and (4) will hold for the most probable number of type 1 bacteria and its posterior probability with q replaced by q_1 and where r is the ratio of the type 1 areas *plus* any area in which their growth might be inhibited to A.

4. The assumption of randomness may have to be modified if, as in some kinds of apparatus, the plate rotates under a rectangular slit and points on it at a distance R from the centre of rotation have a probability of receiving bacteria which varies as $1/R$. All that is required, however, is that each of the areas $a_1, \ldots, a_q$ shall be suitably modified (multiplying by $1/R_0$, where R_0 is the distance of the centre of gravity of the area from the centre of rotation, will be sufficient unless it is a large area).

5. If qr is large it is necessary to calculate $f(m)$ from (4) for quite a number of values of m before obtaining a clear idea of this posterior probability distribution. However, it is possible to represent $f(m)$ approximately, if $qr \geqslant 10$, by a continuous normal distribution with a maximum at n and a standard deviation

$$(qr)^{\frac{1}{2}}/(1-r). \tag{5}$$

For, replacing $\log(m!)$ by $\log\Gamma(1+m)$ and this, in turn, by its asymptotic expansion

$$(m+\tfrac{1}{2})\log m - m + \tfrac{1}{2}\log(2\pi),$$

we find that

$$\log f(m) \doteqdot m\log(mr) - (m-q)\log(m-q) + \tfrac{1}{2}\log[m/(m-q)] + \text{constant}, \tag{6}$$

$$\frac{d\log f(m)}{dm} \doteqdot \log(mr) - \log(m-q) + \frac{1}{2}\left[\frac{1}{m} - \frac{1}{m-q}\right], \tag{7}$$

$$\frac{d^2\log f(m)}{dm^2} \doteqdot \left[\frac{1}{m} - \frac{1}{m-q}\right] + \frac{1}{2}\left[\frac{1}{(m-q)^2} - \frac{1}{m^2}\right]. \tag{8}$$

(7) shows that $f(m)$ has a maximum when

$$m = \frac{q}{1-r} - \frac{1}{2} + O\left(\frac{1}{qr}\right) = m_0, \text{ say.} \tag{9}$$

For this value of m, (8) shows that

$$\frac{d^2\log f(m)}{dm^2} = -\frac{(1-r)^2}{qr} + O\left(\frac{1}{q^3r^3}\right). \tag{10}$$

We now compare $f(m)$ with a distribution

$$g(m) \propto \exp\{-[m - q/(1-r) + \tfrac{1}{2}]^2/[2\sigma^2]\}. \tag{11}$$

(5) is now proved by equating the second derivatives of $\log f(m)$ and $\log g(m)$, provided higher derivatives have only a small effect. To estimate the magnitude of this effect consider the term involving the third derivative in the Taylor series of $\log f(m)$ about $m = m_0$. For $m = m_0 \pm 2(qr)^{\frac{1}{2}}/(1-r)$ we then find that

$$f(m) = \exp\left[-2 \pm \frac{4(1+r)}{3(qr)^{\frac{1}{2}}} + O\left(\frac{1}{qr}\right)\right].$$

An alternative normal curve, for which I am indebted to the referee, may be obtained by using the fact that $f(m)$ is the general term of the negative binomial

$$\left(\frac{1}{1-r} - \frac{r}{1-r}\right)^{-(q+1)}$$

which has a mean $(q+r)/(1-r)$ and a variance $r(q+1)/(1-r)^2$. These may be taken as the mean and variance of the equivalent normal curve and the $\frac{1}{2}$ 'correction for continuity' used (Yates, 1934). The difference between the two normal curves is not significant if $q = O(100)$ which is usually the case.

6. In counting dust particles, the particles are usually graded by the size of their apparent areas. Now smaller particles may be obscured by, or overlap larger particles, or may overlap each other to form apparently larger particles; thus the count of small particles is an underestimate. To calculate a correction to the number of particles q, say, in a particular grade of size, we find the area in which such

particles could be covered by, lie on the top of, or by overlapping form part of, larger particles. Let r be the ratio of this area to the total area in which particles could fall. Then, by the same argument as in Section 2, equation (3) gives the most probable number of particles of the given grade. It is true that this does not tell us how many of the larger particles are really composed of several smaller ones and that the count of these larger particles should be reduced. Mostly such combinations will consist of one small overlapping one large particle, and if the small particle is subtracted from the apparent area the particle will usually remain in the same grade. The reduction due to this effect is thus, in general, unimportant. In any case the theoretical results worked out in this paper should enable experimenters to decide what concentration of dust particles (or bacterial colonies) is permissible to achieve a desired reliability.

REFERENCES

ARMITAGE, P. (1949). *Biometrika*, **36**, 257.
MACK, C. (1948). *Phil. Mag.* **39**, 778.
MACK, C. (1949). *Proc. Camb. Phil. Soc.* **46**, 285.
LIDWELL, O. M. (1948). *Spec. Rep. Ser. Med. Res. Coun., Lond.*, no. 262, pp. 48 and 341.
YATES, F. (1934). *Suppl. J. R. Statist. Soc.* **1**, 217.

Non-normality in two-sample *t*-tests

BY D. G. C. GRONOW, *Institute of Aviation Medicine*

1. In a recent note (1951) by the present author the bias and power of two-sample t-tests was discussed. The notation in the present paper is somewhat different to that employed before, and accordingly the previous discussion is summarized here using the new notation. Given two samples $x_i\,(i = 1, 2, \ldots, n)$ and $x'_j\,(j = 1, 2, \ldots, n')$, two criteria,

$$u = \frac{\bar{x}-\bar{x}'}{\left(\dfrac{n+n'}{nn'}\left[\dfrac{(n-1)\,s^2+(n'-1)\,s'^2}{n+n'-2}\right]\right)^{\frac{1}{2}}} \quad \text{and} \quad v = \frac{\bar{x}-\bar{x}'}{\left(\dfrac{s^2}{n}+\dfrac{s'^2}{n'}\right)^{\frac{1}{2}}},$$

where

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \bar{x}' = \frac{1}{n'}\sum_{j=1}^{n'} x'_j$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2, \quad s'^2 = \frac{1}{n'-1}\sum_{j=1}^{n'}(x'_j-\bar{x}')^2,$$

were considered under the assumption that the populations generating the samples were both normally distributed, but might have different variances. The method used to investigate bias and power was to obtain approximations (in series form) to the moments of u and of v, to approximate to the distributions of these criteria by frequency curves having the same first four moments, and to find the proportion of the frequency curve cut off by a nominal significance level. This procedure worked reasonably well over the range of values considered, but the algebra involved in the calculation of the moments was heavy. It was clear, therefore, that the labour involved in calculating the moments of the criteria under the assumption that the samples had come from a non-normal population would be prohibitive. Accordingly another method was used.

2. David & Johnson (1951) have pointed out that the distribution of any ratio of two random variables is found usually in order to be able to make a probability statement of the form

$$\mathrm{P}\{x/y > L_{\frac{1}{2}\alpha}\} = \tfrac{1}{2}\alpha,$$

where $L_{\frac{1}{2}\alpha}$ is the value of the abscissa of the distribution, the ordinate of which cuts off a proportion $\frac{1}{2}\alpha$ in the tail. This probability statement is the same as

$$\mathrm{P}\{(x-L_{\frac{1}{2}\alpha}y) > 0\} = \tfrac{1}{2}\alpha,$$

and, following Cramér, they pointed out that the distribution of $x-L_{\frac{1}{2}\alpha}y$ could equally well be studied as that of the ratio, provided y is never negative. It will be noted that the criteria u and v can both be expressed in terms of k-statistics as functions such as

$$b(k_1-k'_1)\,(k_2+ak'_2)^{-\frac{1}{2}},$$

where a and b are different for u and for v, except where $n = n'$, when $u \equiv v$. Thus we may make the statement

$$\mathrm{P}\{|\, b(k_1 - k_1')\,(k_2 + ak_2')^{-\frac{1}{2}}\,| > L_{\frac{1}{2}\alpha}\} = \tfrac{1}{2}\alpha,$$

or

$$\mathrm{P}\{b^2(k_1 - k_1')^2 - L^2_{\frac{1}{2}\alpha}(k_2 + ak_2') > 0\} = \alpha.$$

The moments of

$$J = b^2(k_1 - k_1')^2 - L^2_{\frac{1}{2}\alpha}(k_2 + ak_2')$$

can be written down under very general conditions. If δ is the difference between the population means, and if the dashed cumulants relate to the second population, then the first two moments of this quantity are

$$\mathscr{E}(J) = b^2\delta^2 + b^2\left(\frac{\kappa_2}{n} + \frac{\kappa_2'}{n'}\right) - L^2_{\frac{1}{2}\alpha}(\kappa_2 + a\kappa_2'),$$

$$\begin{aligned}\mathscr{E}(J^2) = b^4\delta^4 &+ 2b^2\delta^2\left[3b^2\left(\frac{\kappa_2}{n} + \frac{\kappa_2'}{n'}\right) - L^2_{\frac{1}{2}\alpha}(\kappa_2 + a\kappa_2')\right] \\ &+ 4b^2\delta\left[b^2\left(\frac{\kappa_3}{n^2} - \frac{\kappa_3'}{n'^2}\right) - L^2_{\frac{1}{2}\alpha}\left(\frac{\kappa_3}{n} - \frac{a\kappa_3'}{n'}\right)\right] \\ &+ b^4\left[\frac{\kappa_4}{n^3} + \frac{\kappa_4'}{n'^3} + 3\left(\frac{\kappa_2}{n} + \frac{\kappa_2'}{n'}\right)^2\right] \\ &- 2b^2L^2_{\frac{1}{2}\alpha}\left[\frac{\kappa_4}{n^2} + \frac{a\kappa_4'}{n'^2} + (\kappa_2 + a\kappa_2')\left(\frac{\kappa_2}{n} + \frac{\kappa_2'}{n'}\right)\right] \\ &+ L^4_{\frac{1}{2}\alpha}\left[\frac{\kappa_4}{n} + \frac{a^2\kappa_4'}{n'} + 2\left(\frac{\kappa_2^2}{n-1} + \frac{a^2\kappa_2'^2}{n'-1}\right) + (\kappa_2 + a\kappa_2')^2\right].\end{aligned}$$

The third and fourth moments follow similarly.

3. The moments calculated in §2 are quite general, the only assumption that is made about each population being that its cumulants exist up to any required order. It is possible, however, to simplify the expressions for these calculated moments somewhat if we make a further assumption regarding the functional form of the population sampled. In previous work it has been found that asymmetry in the population generating the samples has little effect on sampling distributions, but that kurtosis may alter nominal significance levels appreciably. We shall assume, therefore, that the parent populations of an investigation may be described by two symmetrical Gram-Charlier Type A series. It will follow that

$$\kappa_3 = \kappa_5 = \kappa_6 = \kappa_7 = 0 \quad \text{and} \quad \kappa_8 = -35\kappa_4^2$$

resulting in considerable simplification of the moments of J. We have that

$$\mu_1'(J) = b^2\delta^2 + b^2\left(\frac{\kappa_2}{n} + \frac{\kappa_2'}{n'}\right) - L^2_{\frac{1}{2}\alpha}(\kappa_2 + a\kappa_2'),$$

$$\begin{aligned}\mu_2(J) = 4b^2\delta^2&\left(\frac{\kappa_2}{n} + \frac{\kappa_2'}{n'}\right) \\ &+ b^4\left(\frac{\kappa_4}{n^3} + \frac{\kappa_4'}{n'^3}\right) - 2b^2L^2_{\frac{1}{2}\alpha}\left(\frac{\kappa_4}{n^2} + \frac{a\kappa_4'}{n'^2}\right) + L^4_{\frac{1}{2}\alpha}\left(\frac{\kappa_4}{n} + \frac{a^2\kappa_4'}{n'}\right) \\ &+ 2b^4\left(\frac{\kappa_2}{n} + \frac{\kappa_2'}{n'}\right)^2 + 2L^4_{\frac{1}{2}\alpha}\left(\frac{\kappa_2^2}{n-1} + \frac{a^2\kappa_2'^2}{n'-1}\right),\end{aligned}$$

with similar, longer expressions for μ_3 and μ_4.

4. Having obtained the moments of J it is necessary to approximate to its distribution by assuming for this distribution a functional form, the parameters of which have been found from the known moments. Preliminary calculations showed that the momental constants β_1 and β_2 would indicate a Pearson Type IV as an approximate graduation. These curves are, however, intractable to handle arithmetically and Johnson (1949) 'unbounded' type curves (S_U) were used instead. Using both tails of the t-distribution, for the nominal test, i.e. a two-tail test, the bias due to kurtosis, in two specific cases, is shown in Tables 1 (*a*) and 1 (*b*). The effect of non-normality is seen to be noticeable for both the u and v criteria, but it seems to be somewhat worse for v than for u when the sample sizes are unequal. This suggests that if the populations may be non-normal as well as have unequal variances, the advantage pointed out by Welch (1937) of using v rather than u when comparing the means of small unequal samples is no longer so clear-cut.

Table 1 (*a*). *Probability, as a percentage, that* $u\,(=v)$ *falls beyond the* 5 % *points of the* t-*distribution with* 18 *degrees of freedom. Case* $n = n' = 10$, $\delta = 0$

(Symmetrical Gram-Charlier populations, $\beta_2 = \beta_2'$.)

κ_2'/κ_2 β_2	1	2	3
2·5	5·33	5·58	5·92
Normal	5·00	5·13	5·30
4·0	5·51	5·74	6·13
5·0	5·94	6·29	7·31

Table 1 (*b*). *Probability, as a percentage, that* u *and* v *fall beyond the* 5 % *points of the* t-*distribution with* 18 *degrees of freedom. Case* $n = 15$, $n' = 5$, $\delta = 0$

(Symmetrical Gram-Charlier populations, $\beta_2 = \beta_2'$).

Criterion	κ_2'/κ_2 β_2	$\frac{1}{3}$	$\frac{1}{2}$	1	2	3
u	2·5	1·56	2·45	5·28	10·34	16·49
	Normal	1·40	2·25	5·00	9·77	13·20
	4·0	1·65	2·58	5·91	10·73	14·59
	5·0	1·83	2·81	6·55	11·14	14·93
v	2·5	5·83	7·14	9·12	11·65	12·74
	Normal	5·47	5·85	6·83	7·99	8·67
	4·0	6·37	6·89	8·76	11·04	12·01
	5·0	7·50	9·11	11·37	13·32	15·23

Table 2. *Effect of non-normality in parent populations on the power of the* u *and* v *tests to detect a difference* δ *between population means* (*using the* 5 % *significance points for* t *with* 18 *degrees of freedom*). *Case* $n = n' = 10$, $\alpha = 0{\cdot}05$

Power of $u\,(=v)$ test.

κ_2'/κ_2	$\dfrac{\delta}{\sqrt{[\frac{1}{2}(\kappa_2+\kappa_2')]}}$ β_2	0	0·5	1·0	1·5	2·0	2·5	3·0
1	2·5	0·053	0·196	0·590	0·893	0·983	0·997	1·000
	Normal	**0·050**	**0·182**	**0·562**	**0·892**	**0·987**	**0·998**	**1·000**
	4·0	0·055	0·205	0·594	0·887	0·980	0·997	1·000
	5·0	0·059	0·216	0·587	0·883	0·978	0·997	1·000
2	2·5	0·056	0·197	0·596	0·894	0·984	0·998	1·000
	Normal	**0·051**	**0·188**	**0·576**	**0·892**	**0·988**	**0·999**	**1·000**
	4·0	0·057	0·207	0·583	0·889	0·980	0·998	1·000
	5·0	0·063	0·222	0·581	0·889	0·977	0·998	1·000
3	2·5	0·059	0·200	0·594	0·889	0·986	0·998	1·000
	Normal	**0·053**	**0·185**	**0·559**	**0·893**	**0·987**	**0·999**	**1·000**
	4·0	0·061	0·212	0·573	0·885	0·980	0·998	1·000
	5·0	0·073	0·224	0·590	0·879	0·978	0·998	1·000

By varying the value of δ, the distance between the population means, we may study the effect of changes in β_2 as well as in the variance ratio κ_2'/κ_2 on the power of the tests, that is to say on the probability of establishing significance. For the case of equal sample sizes, with $n = n' = 10$, Table 2 shows the power of the test for increasing values of the ratio

$$\theta = \delta/\{\tfrac{1}{2}(\kappa_2+\kappa_2')\}^{\frac{1}{2}},$$

which appears to be the appropriate non-central parameter to use. Remembering the differences between the effective significance levels, i.e. between the figures in the column headed $\theta = 0$, it will be seen that in this case of equal samples, the power function is remarkably little modified either by inequality in population variance or departure from normality. This result suggests that provided the two samples are equal or nearly so (as they generally will be in any planned experiment), the standard normal theory power function for the t-test may be used, even when the variances are unequal and β_2 differs from 3.

Thus, for example, in assessing roughly how large samples must be to establish a given difference in population means using the t-test, Pearson & Hartley's (1951, p. 115) chart for $\nu_1 = 1$ could be used with

$$\phi = \tfrac{1}{2}\delta\sqrt{n}/\sigma, \quad \nu_2 = n+n'-2,$$

where the value taken for σ^2 would be a rough estimate of the average population variance, $\frac{1}{2}(\kappa_2+\kappa_2')$, based on past experience.

When the sample sizes are so unequal that the effective significance levels differ substantially from their nominal value, it becomes difficult to interpret the resulting modifications in the power function. Computations have been made for the case $n = 15$, $n' = 5$ and tables of the power function for both u and v have been determined for the same values of β_2 as in Table 2 and for $\kappa_2'/\kappa_2 = \frac{1}{3}, \frac{1}{2}, 1, 2$ and 3. It is hoped to make use of these results in a further communication, but it may be remarked that, as for the significance levels given in Table 1 (*b*) above, the effect of non-equality of variance is far more important than that of kurtosis.

My thanks are due to Dr F. N. David and Professor E. S. Pearson whose suggestions helped considerably in the preparation of this paper.

REFERENCES

David, F. N. & Johnson, N. L. (1951). *Biometrika*, **38**, 43.
Gronow, D. G. C. (1951). *Biometrika*, **38**, 252.
Johnson, N. L. (1949). *Biometrika*, **36**, 149.
Pearson, E. S. & Hartley, H. O. (1951). *Biometrika*, **38**, 112.
Welch, B. L. (1937). *Biometrika*, **29**, 350.

Note on the Poisson Index of Dispersion

By N. KATHIRGAMATAMBY, *University College, London*

1. The distribution of the Poisson Index of Dispersion, and the validity of assuming that it is distributed as χ^2 with degrees of freedom equal to the number of observations minus one, has been discussed at some length by Hoel (1943). He reached the conclusion that the assumption of the χ^2 distribution was valid for a small number of observations provided the Poisson parameter was large, but his method was not altogether satisfactory in that it was based on a comparison of moment ratios, these moments themselves being infinite series. An alternative approach has been to consider the conditional distribution of the Index for a fixed total, say T, of the observed values. Provided T is not too large the probability of every possible partition can be determined and hence the exact overall distribution of the Index obtained by allowing T to vary according to the Poisson law. Cochran (1936) first put forward this method of attack, and further calculations have recently been carried out by Lancaster (1952). Neither dealt fully, however, with the case of more than four observations. We shall concern ourselves here with the case where T may take any value, i.e. with the overall rather than the conditional distribution of the Index. It is possible by means of the rewriting of a probability statement to calculate exact moments and hence by curve fitting to obtain probability integrals which may be compared with those of χ^2. Further, by writing the moments in a general form the power of the Index of Dispersion test can be obtained, when the alternative hypothesis is that the variance is greater than the mean.

2. It is assumed that there are N random independent variables $x_1, x_2, \dots, x_N$ which may take only zero or positive integral values and each of which has the same distribution. If this distribution is Poisson then the Index of Dispersion is written as

$$I = \frac{\sum\limits_{i=1}^{N} (x_i - \bar{x})^2}{\bar{x}},$$

where $\bar{x}$ is the mean, and the probability statement connected with the test of significance takes the form

$$P\{I > \chi^2_{N-1,\,\alpha}\} = \alpha$$

or

$$P\left\{\left(\sum_{i=1}^{N} (x_i - \bar{x})^2 - \chi^2_{N-1,\alpha} \cdot \bar{x}\right) > 0\right\} = \alpha.$$

If k_1 and k_2 denote the first two of Fisher's k-statistics this expression becomes

$$P\{((N-1)\,k_2 - k_1 \chi^2_{N-1,\alpha}) > 0\} = \alpha.$$

Accordingly we focus attention on $\quad S = (N-1)\,k_2 - k_1 \chi^2_{N-1,\alpha},$

the moments of which can be written down exactly. Assume, for the time being, that the cumulants of the parent population are $\kappa_1, \kappa_2, \kappa_3, \dots$. We note that both k_1 and κ_1 can never be negative. Easy algebra gives

$$\begin{aligned}
\mathscr{E}(S) &= (N-1)\,\kappa_2 - \kappa_1 \chi^2_{N-1,\alpha},\\
\mu_2(S) &= (N-1)^2\left[\frac{\kappa_4}{N} + \frac{2\kappa_2^2}{N-1}\right] - 2(N-1)\,\chi^2_{N-1,\alpha}\frac{\kappa_3}{N} + \chi^4_{N-1,\alpha}\frac{\kappa_2}{N},\\
\mu_3(S) &= (N-1)^3\left[\frac{\kappa_6}{N^2} + \frac{12\kappa_4\kappa_2}{N(N-1)} + \frac{4(N-2)}{N(N-1)^2}\kappa_3^2 + \frac{8\kappa_2^3}{(N-1)^2}\right]\\
&\quad - 3(N-1)^2\chi^2_{N-1,\alpha}\left[\frac{\kappa_5}{N^2} + \frac{4\kappa_3\kappa_2}{N(N-1)}\right] + 3(N-1)\,\chi^4_{N-1,\alpha}\frac{\kappa_4}{N^2} - \chi^6_{N-1,\alpha}\frac{\kappa_3}{N^2},\\
\kappa_4(S) &= (N-1)^4\left[\frac{\kappa_8}{N^3} + \frac{24\kappa_6\kappa_2}{N^2(N-1)} + \frac{32(N-2)}{N^2(N-1)^2}\kappa_5\kappa_3 + \frac{8(4N^2-9N+6)}{N^2(N-1)^3}\kappa_4^2\right.\\
&\quad \left. + \frac{144\kappa_4\kappa_2^2}{N(N-1)^2} + \frac{96(N-2)}{N(N-1)^3}\kappa_3^2\kappa_2 + \frac{48\kappa_2^4}{(N-1)^3}\right]\\
&\quad - 4(N-1)^3\chi^2_{N-1,\alpha}\left[\frac{\kappa_7}{N^3} + \frac{12\kappa_5\kappa_2}{N^2(N-1)} + \frac{12\kappa_4\kappa_3}{N^2(N-1)} + \frac{8(N-2)}{N^2(N-1)^2}\kappa_4\kappa_3 + \frac{24\kappa_3\kappa_2^2}{N(N-1)^2}\right]\\
&\quad + 6(N-1)^2\chi^4_{N-1,\alpha}\left[\frac{\kappa_6}{N^3} + \frac{4\kappa_4\kappa_2}{N^2(N-1)} + \frac{4\kappa_3^2}{N^2(N-1)}\right]\\
&\quad - 4(N-1)\,\chi^6_{N-1,\alpha}\frac{\kappa_5}{N^3} + \chi^8_{N-1,\alpha}\frac{\kappa_4}{N^3}.
\end{aligned}$$

When the parent population is Poisson

$$\kappa_r = \lambda \quad \text{for} \quad r = 1, 2, \dots,$$

and the moments of S reduce to

$$\begin{aligned}
\mathscr{E}(S) &= \lambda[N - 1 - \chi^2_{N-1,\alpha}],\\
\mu_2(S) &= \lambda\left[\frac{(N-1)^2}{N} - \frac{2(N-1)}{N}\chi^2_{N-1,\alpha} + \frac{1}{N}\chi^4_{N-1,\alpha}\right] + 2\lambda^2(N-1),\\
\mu_3(S) &= \lambda\left[\frac{(N-1)^3}{N^2} - \frac{3(N-1)^2}{N^2}\chi^2_{N-1,\alpha} + \frac{3(N-1)}{N^2}\chi^4_{N-1,\alpha} - \frac{1}{N^2}\chi^6_{N-1,\alpha}\right]\\
&\quad + \lambda^2\left[\frac{4(N-1)(4N-5)}{N} - \frac{12(N-1)}{N}\chi^2_{N-1,\alpha}\right] + 8\lambda^3(N-1),\\
\kappa_4(S) &= \frac{\lambda}{N^3}[(N-1)^4 - 4(N-1)^3\chi^2_{N-1,\alpha} + 6(N-1)^2\chi^4_{N-1,\alpha} - 4(N-1)\,\chi^6_{N-1,\alpha} + \chi^8_{N-1,\alpha}]\\
&\quad + \frac{\lambda^2}{N^2}[8(N-1)[11N^2 - 27N + 17] - 32(N-1)(4N-5)\,\chi^2_{N-1,\alpha} + 48(N-1)\,\chi^4_{N-1,\alpha}]\\
&\quad + \frac{\lambda^3}{N}[48(N-1)(5N-7) - 96(N-1)\,\chi^2_{N-1,\alpha}] + 48\lambda^4(N-1).
\end{aligned}$$

3. Having obtained the moments it is necessary to decide on a suitable functional form with which to approximate to the distribution. Preliminary calculations showed that the β_1 and β_2 of the distribution of S lay in the Type IV area of the Pearson system. While it is natural to fit a Pearson curve, since so many have occurred naturally in sampling theory, the Type IV is intractable to handle. Moreover recent investigations have demonstrated the closeness of the probability integrals of the Type IV and the Johnson S_u curves with the same first four moments, and the latter system is considerably easier to compute. In the main therefore either the Johnson S_u or the log-normal curves were used to estimate the probability integral from the moments. In two cases Pearson Type IV's were fitted as a check on the Johnson system and in nine cases Edgeworth's form of the Gram-Charlier series. The difference between the probability integrals estimated by all three curves is not large. The results are given in Table 1. These estimated probabilities should be equal to a value of $\alpha = 0{\cdot}05$ if the χ^2 distribution is a satisfactory approximation to the true distribution of the Index of Dispersion. The interdependence of N and λ will be noticed from the table. For either N large or λ large, or both large, as Hoel noticed, the χ^2 distribution is a reasonably good approximation, and in fact, there appears to be little risk of error for N as small as 6 provided λ is greater than or equal to unity. Generally the larger the sample size the smaller the acceptable value of λ, but this rule must not be pushed to extremes. I, in practice, is discontinuous and may only take a finite set of values for given λ and N, a set which will decrease rapidly with decreasing λ. In fact for λ small discontinuities are more likely to be a source of error than assuming the distribution of I is χ^2.

Table 1. *Estimated probability integrals corresponding to a nominal* χ^2 *level of* 0·05

λ / N	1			5		10	
	S_u or S_L	Gr.-Ch.	Type IV	S_u or S_L	Gr.-Ch.	S_u or S_L	Gr.-Ch.
201	0·051	0·052	0·052	0·050	0·050	0·050	0·050
101	0·051	0·052	0·051	0·050	0·050	—	—
51	0·051	—	—	0·050	—	0·050	—
11	0·052	—	—	0·048	—	0·049	—
6	0·055	0·050	—	0·048	0·053	0·048	0·052

One further interesting point is suggested by these results. For any given sample size it appears that for λ small the first kind of error is greater than that allowed for by the χ^2 distribution. This in judging significance from χ^2 tables we shall tend to overestimate significance for λ small and a low nominal significance level should therefore be chosen.

Table 2. *Power of test when the alternative hypothesis is a compound Poisson.*
(*Nominal significance level* = 0·05)

Type of series	λ / N	$\kappa_2/\kappa_1 = \frac{3}{2}$			$\kappa_2/\kappa_1 = 2$		
		1	5	10	1	5	10
Thomas	51	0·635	0·667	0·669	0·956	0·960	—
Neyman		0·665	0·666	0·669	—	—	—
Neg. binomial		0·606	0·659	0·666	0·898	0·949	0·955
Thomas	101	0·870	0·888	0·890	0·999	1·000	0·999
Neyman		0·865	0·887	0·890	0·999	0·999	—
Neg. binomial		0·834	0·880	0·886	0·988	0·997	0·999

4. The Index of Dispersion is commonly used to detect departures from the Poisson distribution in the shape of the variance being larger than the mean. There are many population forms which can be chosen to describe this state of affairs; we have confined ourselves to such distributions as can be described by a two-parameter compound Poisson series. For the sake of example we considered three types of such a distribution, the Thomas, the Neyman, and the negative binomial. The parameters of these series can be chosen so that they all have the same mean and variance, but they will have different higher cumulants. Substitution of the cumulants of any one of these in the general expression for the moments of S enables us to obtain an approximate distribution from which we can find $P\{S > 0\}$ and hence the power of the index of dispersion test with regard to an alternative of the selected compound Poisson type. The results of such calculations are given in Table II.

The power of the test appears to depend but little on the value of λ (or κ_1) but increases, as is expected, with increase in sample size and with the value of κ_2/κ_1. The differences in the power of the test with regard to the three alternative series are not large.

REFERENCES

COCHRAN, W. G. (1936). *Ann. Eugen., Lond.*, **7**, 207.
HOEL, P. G. (1943). *Ann. Math. Statist.* **14**, 155.
LANCASTER, H. O. (1952). *Biometrika*, **39**, 419.

On an extension of Geary's theorem

BY R. G. LAHA, *Statistical Laboratory, Calcutta*

R. C. Geary (1936) has shown that if the sample mean is distributed independently of the sample variance, then the population should be normal. He assumed that the moments of all order in the population should exist. Later on, E. Lukacs (1942) proved the same theorem, assuming only that the population should possess a finite variance.

Basu & Laha (1952) extended Geary's theorem by proving that if the sample mean is distributed independently of any sample k-statistic (as defined by R. A. Fisher), say k_r $(r \geqslant 2)$, then the population is normal.

Here another extension of Geary's theorem is given.

THEOREM. Let $x_1, x_2, \ldots, x_n$ be identically distributed, independent random variables with a finite variance σ^2.

If now the conditional expectation of any unbiased quadratic estimate of $c\sigma^2$ $(c \neq 0)$ for given $x_1 + x_2 + \ldots + x_n$ does not involve the latter, then each x should be normally distributed.

Proof. Let $Q = \sum_{i,j=1}^{n} a_{ij} x_i x_j$ be any unbiased quadratic estimate of $c\sigma^2$ $(c \neq 0)$.

Then, from the definition of unbiasedness, we have

$$E(Q) = c\sigma^2,$$

or

$$\sigma^2 \sum_{i=1}^{n} a_{ii} + m^2 \sum_{i,j=1}^{n} a_{ij} = c\sigma^2, \quad \text{where} \quad m = E(x),$$

or

$$\sum_{i=1}^{n} a_{ii} = c \quad \text{and} \quad \sum_{i,j=1}^{n} a_{ij} = 0. \tag{1·1}$$

Now, from the condition of the theorem, it follows evidently that

$$\begin{aligned} E\{Q\, e^{it\Sigma x}\} &= E(Q)\, E(e^{it\Sigma x}) \\ &= c\sigma^2 \{\phi(t)\}^n, \end{aligned} \tag{1·2}$$

where $\phi(t) = E(e^{itx})$ stands for the characteristic function of the distribution of x.

Now simplifying the left-hand side of (1·2), it can be shown easily that (1·2) reduces to the form

$$\left(\frac{d^2\phi}{dt^2}\Big/\phi\right) \sum_{i=1}^{n} a_{ii} + \left(\frac{d\phi}{dt}\Big/\phi\right)^2 \sum_{i \neq j} a_{ij} = -c\sigma^2. \tag{1·3}$$

Let us now write $\psi(t) = \log \phi(t)$, where $\psi(t)$ represents the cumulant-generating function of the distribution of x, then

$$\frac{d\psi}{dt} = \frac{d\phi}{dt}\Big/\phi \quad \text{and} \quad \frac{d^2\psi}{dt^2} = \frac{d^2\phi}{dt^2}\Big/\phi - \left(\frac{d\phi}{dt}\Big/\phi\right)^2. \tag{1·4}$$

Substituting (1·4) in (1·3) we have

$$\frac{d^2\psi}{dt^2}\sum_{i=1}^{n} a_{ii} + \left(\frac{d\psi}{dt}\right)^2 \sum_{i,j=1}^{n} a_{ij} = -c\sigma^2. \qquad (1\cdot5)$$

Now, from the relations in (1·1), (1·5) at once gives

$$\frac{d^2\psi}{dt^2} = -\sigma^2, \qquad (1\cdot6)$$

that is, $\psi(t)$ is a polynomial of degree 2, which shows evidently that x should be normally distributed. It is interesting to note here that we need not assume the complete stochastic independence of Σx and Q.

From this theorem and Craig's theorem (1943), it follows evidently that if the sum $x_1 + x_2 + \ldots + x_n$ is distributed independently of any quadratic form $\sum_{i,j=1}^{n} a_{ij}x_ix_j$ satisfying the relations

$$\text{(i)}\quad \sum_{i=1}^{n} a_{ii} \neq 0 \quad \text{and} \quad \text{(ii)}\quad \sum_{i,j=1}^{n} a_{ij} = 0$$

then

$$\epsilon A = 0,$$

where A is the matrix of the quadratic form, and $\epsilon = (1, 1, \ldots, 1)$.

REFERENCES

Basu, D. & Laha, R. G. (1952). On some characterisations of the normal distribution. *Sankhyā* (in the Press).

Craig, A. T. (1943). Note on the independence of certain quadratic forms. *Ann. Math. Statist.* **14**, 195–7.

Geary, R. C. (1936). The distribution of the Student's ratio for the non-normal samples. *Suppl. J. R. Statist. Soc.* **3**, 178–84.

Lukacs, E. (1942). A characterisation of the normal distribution. *Ann. Math. Statist.* **13**, 91–3.

The Doolittle method and the fitting of polynomials to weighted data

By P. G. GUEST, *University of Sydney, Australia*

A modification of the standard Doolittle scheme of solving normal equations to include the calculation of orthogonal polynomials and standard errors has recently been described (Guest, 1950). In the present communication the case in which the observations are weighted and factorial moments are employed is considered. This has been discussed by Aitken, Fisher & Kimball, and it is the purpose of this note to correlate and extend their treatments.

The least-squares curve $u_p(x)$ of degree p which fits the n observations $y(x_i)$, of weight $w(x_i)$, is obtained from the condition that

$$\sum_{i=1}^{n} w(x_i)\{y(x_i) - u_p(x_i)\}^2$$

should be a minimum. We consider the case in which it is desired to express $u_p(x)$ as a linear function of polynomials $f_j(x)$ of degree j in x, j taking values from 0 to p. Then

$$u_p(x) = \sum_{j=0}^{p} b_{pj} f_j(x), \qquad (1)$$

where the b_{pj} are given by the normal equations

$$\sum_i w(x_i)\{y(x_i) - \sum_j b_{pj} f_j(x_i)\} f_k(x_i) = 0 \quad (k = 0 \text{ to } p). \qquad (2)$$

If we write

$$\sum_i w(x_i)\, y(x_i)\, f_k(x_i) = M_k, \qquad (3)$$

$$\sum_i w(x_i)\, f_j(x_i)\, f_k(x_i) = \phi_{jk}, \qquad (4)$$

then the normal equations become

$$\sum_{j=0}^{p} b_{pj}\phi_{jk} = M_k. \qquad (5)$$

The quantities ϕ_{jk} form a symmetric matrix, and the standard Doolittle or abbreviated Doolittle method of solution (Dwyer, 1941) may be employed.

We can introduce orthogonal polynomials $T_j(x)$, of degree j in x, satisfying the equations

$$\sum_i w(x_i)\, T_j(x_i)\, T_k(x_i) = 0 \quad (j \neq k). \tag{6}$$

These equations determine the orthogonal polynomials except for an arbitrary factor. This factor is conveniently chosen so that the leading coefficient of $T_j(x)$ and $f_j(x)$ are identical. Thus

$$T_j(x) = f_j(x) + \sum_{k=0}^{j-1} \alpha_{jk} T_k(x), \tag{7}$$

where, using (6),

$$\alpha_{jk} = -\sum_i w(x_i)\, f_j(x_i)\, T_k(x_i) \Big/ \sum_i w(x_i)\, T_k^2(x_i). \tag{8}$$

Writing

$$S_{jk} = \Sigma\, w(x_i)\, f_j(x_i) T_k(x_i), \tag{9}$$

$$\alpha_{jk} = -S_{jk}/S_{kk} \tag{10}$$

and

$$\begin{aligned} S_{jk} &= \sum_i w(x_i)\, f_j(x_i)\, \{f_k(x_i) + \alpha_{k,k-1} T_{k-1}(x_i) + \ldots + \alpha_{k0} T_0(x_i)\} \\ &= \phi_{jk} + \sum_{m=0}^{k-1} \alpha_{km} S_{jm}. \end{aligned} \tag{11}$$

The quantities S_{jk}, α_{jk} can thus be built up in succession from the given ϕ_{jk} and the previously calculated values S_{jm}, α_{km} $(m < k)$.

If $u_p(x)$ is expanded in terms of the polynomials $T_j(x)$,

$$u_p(x) = \sum_{j=0}^{p} a_j T_j(x), \tag{12}$$

then the least-squares condition leads to the equations

$$a_j = M_j'/S_{jj}, \tag{13}$$

where

$$\begin{aligned} M_j' &= \Sigma\, w(x_i)\, y(x_i)\, T_j(x_i) \\ &= M_j + \sum_{k=0}^{j-1} \alpha_{jk} M_k'. \end{aligned} \tag{14}$$

The author has pointed out that the quantities S_{jk}, α_{jk}, M_j', a_j are obtained directly in the forward section of the standard Doolittle method of solving the normal equations. For example, in Dwyer's notation, $S_{jk} = \phi_{jk.12..(k-1)}$ and $M_j' = M_{j.12..(j-1)}$. Further, if coefficients β_{jk} are introduced such that

$$T_j(x) = \sum_{k=0}^{j} \beta_{jk} f_k(x), \tag{15}$$

then

$$\beta_{jk} = \alpha_{jk} + \sum_{m=k+1}^{j-1} \alpha_{jm} \beta_{mk}, \tag{16}$$

and the β_{jk} can be built up in the Doolittle table along with the α_{jk}.

The coefficients b_{pj} can be written as

$$b_{pj} = a_j + \sum_{j+1}^{p} \alpha_{kj} b_{pk} \tag{17}$$

or

$$b_{pj} = a_j + \sum_{j+1}^{p} \beta_{kj} a_k. \tag{18}$$

Equation (17) is used in the backwards section of the standard Doolittle solution. However, the expansion in terms of the β_{kj} has the advantage that the coefficients b_{rj} in the least-squares polynomials of degree r $(<p)$ are obtained by merely omitting the terms $\beta_{kj} a_k$ in (18) for which $k > r$. The use of the quantities β in the estimation of standard errors has been described in the earlier paper.

The polynomials $f_j(x)$ are commonly x^j or $\dbinom{x}{j}$. Whether power moments or reduced factorial moments are used depends both on the type of machine available and on the form in which the least-squares curve is required. If power moments are used, then $\phi_{jk} = \Sigma w(x)\, x^{j+k}$. Power moments are always used when the data do not occur at equally spaced intervals.

When reduced factorial moments are used, $\phi_{jk} = \Sigma w(x)\binom{x}{j}\binom{x}{k}$, which is not directly calculable by repeated summation. However, ϕ_{jk} can be expanded as a linear function of the $\Sigma w(x)\binom{x}{m}$. In fact

$$\phi_{jk} = \sum_{m=0}^{k}\binom{k}{m}\binom{j+k-m}{k}\sum_x w(x)\binom{x}{j+k-m}. \tag{19}$$

In the method of calculation described by Fisher, equation (5) is used, with the ϕ_{jk} calculated from equation (19).

Aitken, however, uses for the functions $f_j(x)$ the unreduced factorials $x^{(j)} = x(x-1)\dots(x-j+1)$, and by linear combination converts the equations (5) into the equivalent set

$$\sum_{j=0}^{p} b_{pj}\sum_i w(x_i)\,(x_i+k)^{(j+k)} = \sum_i w(x_i)\,y(x_i)\,(x_i+k)^{(k)}. \tag{20}$$

The sums and moments occurring in these equations can be obtained directly by repeated summation. But the matrix of the coefficients of b_{pj} is now non-symmetric, and the Doolittle method of solution cannot be used. For this reason Fisher's method of obtaining the normal equations is to be preferred, and these equations should be solved by the Doolittle technique.

REFERENCES

AITKEN, A. C. (1933). *Proc. Roy. Soc. Edinb.* **54**, 1.

DWYER, P. S. (1941). *Psychometrika*, **6**, 101.

FISHER, R. A. (1946). *Statistical Methods for Research Workers*, 10th ed., p. 166. Edinburgh: Oliver and Boyd.

GUEST, P. G. (1950). *Phil. Mag.* [7], **41**, 124.

KIMBALL, B. F. (1940). *Ann. Math. Statist.* **11**, 348.

A simple method of deriving best critical regions similar to the sample space in tests of an important class of composite hypotheses

BY K. S. RAO,* *Department of Statistics, University of Bombay*

1. *Introduction.* The theory of testing of hypotheses, developed by Neyman & Pearson in a series of articles commencing from 1928, is an integral part of the syllabus in mathematical statistics at various universities. In teaching the subject, the author of the present paper found that it was possible to shorten the rather formal step which Neyman & Pearson (1933) took in their paper 'On the problem of the most efficient tests of statistical hypotheses', in finding the test statistic to be used for a uniformly most powerful test of a composite hypothesis regarding a normal universe. By using a simple lemma given below, a great deal of formal algebra of the original paper can be saved.

2. *The lemma.* The necessary and sufficient condition that the variates X and Y may be independently distributed as $\chi^2\sigma^2$-variates, with m and n as their respective degrees of freedom, is that the variates U and V defined by

$$U = \frac{n}{m}\frac{X}{Y}, \quad V = X+Y, \tag{1}$$

are independently distributed, respectively, as an F-variate with m and n degrees of freedom and a $\chi^2\sigma^2$-variate with $m+n$ degrees of freedom. In other words, the sum of two independent $\chi^2\sigma^2$-variates is independent of their ratio.

The proof† is easy and therefore omitted. An extension of this lemma is that, if $X_1, \dots, X_r$ are independent $\chi^2\sigma^2$-variates with $m_1, \dots, m_r$ degrees of freedom, then the r statistics

$$\frac{m_1 X_1}{m_2 X_2}, \quad \frac{m_1+m_2}{m_3}\frac{X_3}{X_1+X_2}, \quad \dots, \quad \frac{\sum_1^{r-1} m_s}{m_r}\frac{X_r}{\sum_{s=1}^{r-1} X_s}, \quad X_1+\dots+X_r, \tag{2}$$

are, respectively, $r-1$ independent F-variates and a $\chi^2\sigma^2$-variate and the converse. Multivariate extensions of the above result and their applications are given by K. S. Rao (1951).

---

* Nuffield Foundation Dominion Fellow, 1952–3.

† It will be found in a number of text-books on statistics.

3. *Illustration of the method.* The extent of simplification that results by using the above lemma is illustrated by applying it to Example 10 of Neyman & Pearson (1933, pp. 328–32). The method is general and applicable to similar situations. In what follows, the numbers given in bold type after the serial number of an equation refer to the number of the equation in Neyman & Pearson's paper. This procedure, it is hoped, will bring out clearly the simplification made.

The test is that for the significance of the difference between two variances. Neyman & Pearson consider that two samples,

$$\text{(1) } \Sigma_1 \text{ of size } n_1, \quad \text{mean} = \bar{x}_1, \quad \text{standard deviation} = s_1,$$

$$\text{(2) } \Sigma_2 \text{ of size } n_2, \quad \text{mean} = \bar{x}_2, \quad \text{standard deviation} = s_2,$$

have been drawn at random from *some* normal populations. If this is so, the most general probability law for the observed event may be written

$$p(x_1, \ldots, x_{n_1}; x_{n_1+1}, \ldots, x_N) = \left(\frac{1}{\sqrt{(2\pi)}}\right)^N \frac{1}{\sigma_1^{n_1}\sigma_2^{n_2}} \exp\left[-n_1\frac{(\bar{x}_1-a_1)^2+s_1^2}{2\sigma_1^2} - n_2\frac{(\bar{x}_2-a_2)^2+s_2^2}{2\sigma_2^2}\right], \quad (3, \mathbf{159})$$

where $n_1+n_2 = N$, and a_1, σ_1 are the mean and standard deviation of the first, and a_2, σ_2 of the second sampled population.

The admissible simple hypotheses include pairs of sampled populations for which a_1, a_2, $\sigma_1 > 0$, $\sigma_2 > 0$ may have any values whatever. H_0' is the composite hypothesis that $\sigma_1 = \sigma_2$. This is a test for significance of the difference between two variances in two independent samples. The parameters may be defined as follows:

For a simple alternative H_t:

$$\alpha_t^{(1)} = a_1, \quad \alpha_t^{(2)} = a_2 - a_1 = b_1, \quad \alpha_t^{(3)} = \sigma_1, \quad \alpha_t^{(4)} = \theta_t = \sigma_2/\sigma_1. \quad (4, \mathbf{160})$$

For the hypothesis to be tested, H_0':

$$\alpha^{(1)} = a, \quad \alpha^{(2)} = b, \quad \alpha^{(3)} = \sigma, \quad \alpha_0^{(4)} = 1. \quad (5, \mathbf{161})$$

After showing that the conditions required by their theory are satisfied, they proceed to construct the best critical region similar to the sample space.

In the process they obtain the surfaces

$$\phi_1 = \bar{x}_1 = \text{const.}, \quad (6, \mathbf{171})$$

$$\phi_2 = \bar{x}_2 = \text{const.}, \quad (7, \mathbf{172})$$

$$\phi_3 = \frac{1}{N}(n_1s_1^2 + n_2s_2^2) = s_a^2 = \text{const.} \quad (8, \mathbf{173})$$

The element $w_0(\phi_1, \phi_2, \phi_3)$ is the part of $W(\phi_1, \phi_2, \phi_3)$ within which

$$p_t \geq k(\bar{x}_1, \bar{x}_2, s_a)\, p_0, \quad (9, \mathbf{174})$$

the value of k being determined for each system of values of $\bar{x}_1$, $\bar{x}_2$, s_a, so that

$$P_0(w_0(\phi_1, \phi_2, \phi_3)) = \epsilon P_0(W(\phi_1, \phi_2, \phi_3)). \quad (10, \mathbf{175})$$

The best critical region similar to the sample space is obtained by integrating the element $w_0(\phi_1, \phi_2, \phi_3)$ with respect to ϕ_1, ϕ_2, ϕ_3 over their ranges of possible values. By using condition (**174**) they derive that, for

(*a*) $\theta = \sigma_2/\sigma_1 > 1$: the 'best critical region' (B.C.R.) will be defined by

$$s_2^2 \geq k_1(\bar{x}_1, \bar{x}_2, s_a^2), \quad (11, \mathbf{178})$$

and for

(*b*) $\theta = \sigma_2/\sigma_1 < 1$: the B.C.R. will be defined by

$$s_2^2 \leq k_2(\bar{x}_1, \bar{x}_2, s_a^2). \quad (12, \mathbf{179})$$

After devoting the whole of p. 331 and part of p. 332 to formal algebra, they obtain the best critical regions as follows:

(*a*) For alternatives $\sigma_2 > \sigma_1$:

$$u = \frac{n_2s_2^2}{n_1s_1^2 + n_2s_2^2} \geq u_0. \quad (13, \mathbf{194})$$

(*b*) For alternatives $\sigma_2 < \sigma_1$:

$$u = \frac{n_2s_2^2}{n_1s_1^2 + n_2s_2^2} \leq u_0'. \quad (14, \mathbf{195})$$

However, the lemma quoted above makes the transition from (**178**) to (**194**) very simple as follows:

Dividing both sides of (**178**) by the positive quantity, $(n_1s_1^2 + n_2s_2^2)/n_2 = Ns_a^2/n_2$ in case of (*a*), the B.C.R. will be defined by

$$u = \frac{n_2s_2^2}{n_1s_1^2 + n_2s_2^2} \geq k_3(\bar{x}_1, \bar{x}_2, s_a^2). \quad (15)$$

The left-hand side of the inequality is a function of the ratio of two independent $\chi^2\sigma^2$-variates, $n_1 s_1^2$ and $n_2 s_2^2$ which are independent of $\bar{x}_1$ and $\bar{x}_2$. From the lemma it follows that the left-hand side is statistically independent of $\bar{x}_1$, $\bar{x}_2$ and s_a^2. Hence any percentage point of the statistic u on an intersection surface given by **(171)**, **(172)** and **(173)**, is the same for all such surfaces and hence the B.C.R. is given by **(194)**.

REFERENCES

NEYMAN, J. & PEARSON, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans.* A, **231**, 289–337.

RAO, K. S. (1951). On the mutual independence of a set of Hotelling's T^2's derivable from a sample of size n from a k-variate normal population. *Proc. Int. Statist. Conf., India,*

REVIEWS

Econometrics. By GERHARD TINTNER. New York: John Wiley and Sons. 1952. $5.75.

Quantitative economics has three main branches. First there is the expression of the laws of economic behaviour in a form which permits of empirical application. Secondly, there is the problem, often of great practical difficulty, of extracting from the observed workings of the economic system the data by which the magnitudes entering into these laws may be estimated. Finally, the estimated magnitudes must be applied to the theoretical laws in order to calculate parameters and test hypotheses concerning them. Strictly speaking, all three branches have their place in econometrics, but the usage of the term has tended to become restricted to the last of the three; it is with this branch of the subject that Dr Tintner's book is wholly concerned.

Earlier books on econometrics have dealt mainly with the economic background or with special parts of the field, and this is the first attempt that has been made at a systematic exposition of the statistical methodology of the subject. The book is a compendium of known procedures rather than a research monograph, and thus will be of great value to the practical worker although containing little that will be new to the mathematical statistician.

Part I gives a non-technical introduction and has the great merit of containing a large number of numerical illustrations of the application of the methods to a wide variety of economic problems. This part of the book, with its lavish supply of references, furnishes us incidentally with a valuable historical survey. One confusing feature is the treatment of linear regression theory which is introduced in terms of the linear relation $Y = \alpha + \beta X$. Presumably one should regard this as a functional relation which holds exactly. However, we have next 'the empirical regression equation' $Y' = a + bX$ (1) which 'will not hold exactly'. But we are not told what a and b are and why they differ from α and β. They cannot be the sample estimators, since these are denoted later by a^* and b^* and the residuals $Y - Y'$ are assumed to be independent; on the other hand, they must be random variables since the residuals are random variables. It would surely have been much clearer to have written down the model explicitly as $Y = \alpha + \beta X + \epsilon$ in the first place, any relevant assumptions being made about ϵ. Once this hurdle is over, the exposition is sound and straightforward.

Part II, dealing with multivariate analysis, takes the development into rather deeper levels of statistical methodology, and it is here that one begins to wonder how far the data can support the rather elaborate analyses that are made of them. Throughout the book the author is at some pains to point out the defects of his data, but this cannot be done in detail, and the mere fact that such analyses are made may beguile the uncritical. Much use is made here of models in which the relations are assumed to hold exactly between variables the observations of which are subject to error. Apart from obvious theoretical objections, the use of such models requires a knowledge of the variance matrix of the errors, and this will not be available in practice. The way round this difficulty suggested by the author, namely, to assume that the systematic parts of the variables have smooth time paths and to extract the error components by the variate difference method, does not seem very convincing.

Part III gives a useful survey of the special problems that arise owing to the fact that most economic data occur in the form of time series. Particularly helpful is the stress laid on the analysis of regression data with serially correlated errors.

J. DURBIN

Quality Control and Industrial Statistics. By ACHESON J. DUNCAN. Chicago: Richard D. Irwin, Inc. 1952. xxvii + 663 pp. Price $9.

The book begins with a historical review of the industrial applications of statistics in U.S. and Great Britain and with some quotations from the *Wall Street Journal* and the *New York Times* to testify to the commercial value of the methods. Part I of the book then deals with the fundamental notions of probability, frequency distribution and sampling distribution. Part II is about single sample, double sample and sequential acceptance schemes for proportion defective, and Part III is on single sample schemes for continuous variables. Part IV describes the theory of control charts, and Part V deals with statistical methods likely to be useful in industrial research, including contingency tables, analysis of variance, regression and correlation and experimental design. Proofs that can be obtained by elementary algebra are contained in appendices, and there is a comprehensive set of statistical tables and glossaries of symbols and special technical terms. There are over 300 exercises for the reader, many of them containing sets of data from the American literature on quality control. The book as a whole is copiously illustrated with worked examples, tables, diagrams and nomograms.

Part II on acceptance sampling for proportion defective suffers from the failure to emphasize the fundamental distinction between rectifying and non-rectifying inspection and from not stressing the rôle of cost in determining the amount and type of sampling. The result is a clear account of the main types of scheme but, apart from one short section, no indication is given of which type should be used in given circumstances. A whole chapter in this Part is devoted to a detailed description of the sampling schemes used by the U.S. Department of Defence.

Part IV on control charts follows the usual American practice of setting control limits three standard deviations from the mean even when sampling binomial and Poisson distributions. An unusual feature of the treatment is that the operating characteristics of the charts are always given. The account of 'compressed-limit gauges' (p. 296) is rather misleading, the important paper of W. L. Stevens (*J. R. Statist. Soc.* B, **10** (1948), 54) having been overlooked. The use of control charts to help to improve processes so that they become 'in control' is clearly explained but the inexperienced reader might get the erroneous idea that the state of 'control' is easily attained.

Part V includes, besides the conventional statistical methods, some recent developments not normally appearing in text-books. Examples are the use of short-cut methods of analysis of variance based on range, J. W. Tukey's gap and straggler tests for comparing individual means in analysis of variance, and some interesting work by F. E. Grubbs on comparing the precision of different observers or instruments from results arranged in a two-way table.

A surprising omission is that there is no adequate account of how to obtain samples. There is a brief discussion of the general principles of sampling, but the example given is highly simplified. Some description of special methods of sampling powders, liquids, fibres, etc., would have been very valuable.

According to the author's preface, the book is intended both as a text-book for engineering and business students and as a reference book for quality control engineers and industrial research workers. Engineering and business students in this country, who receive at most a short course on statistics, are unlikely to find this very detailed and lengthy account of use. The excellent collection of exercises will, however, be very valuable to anyone planning a course of lectures on statistics for engineers, and the book can be strongly recommended on this account. Because it is reasonably comprehensive, the book can also be recommended to technologists using quality control and acceptance sampling.

D. R. COX

CORRIGENDA

(1) TABLES OF PERCENTAGE POINTS OF THE 'STUDENTIZED' RANGE

(*Biometrika* (1952), **39**, 192)

I very much regret to have to report inaccuracies in the lines $\nu = 5$ and 7 of the revised tables of this function. These lines were obtained, at the last stage of the computational procedure, by interpolation between neighbouring values of ν. Through an oversight on my part the convergence of the interpolation formula was not checked, nor were the four percentage points of q for $n = 2$, $\nu = 5, 7$ compared with the corresponding percentage points of $\sqrt{2}t$.*

Table of the upper percentage points of the 'Studentized' range, q

ν / n	5% points		1% points	
	5	7	5	7
2	3·64	3·34	5·70	4·95
3	4·60	4·16	6·97	5·92
4	5·22	4·68	7·80	6·54
5	5·67	5·06	8·42	7·01
6	6·03	5·36	8·91	7·37
7	6·33	5·61	9·32	7·68
8	6·58	5·82	9·67	7·94
9	6·80	6·00	9·97	8·17
10	6·99	6·16	10·24	8·37
11	7·17	6·30	10·48	8·55
12	7·32	6·43	10·70	8·71
13	7·47	6·55	10·89	8·86
14	7·60	6·66	11·08	9·00
15	7·72	6·76	11·24	9·12
16	7·83	6·85	11·40	9·24
17	7·93	6·94	11·55	9·35
18	8·03	7·02	11·68	9·46
19	8·12	7·09	11·81	9·55
20	8·21	7·17	11·93	9·65

The realization of this omission made it necessary to develop a completely different interpolation formula, and it was noticed that the ratio q is almost proportional to $\sqrt{F(n-1, \nu)}$, where $F(n-1, \nu)$ is the corresponding percentage point of the variance ratio F based on (i) $n-1$ degrees of freedom for the numerator mean square, where n is the size of the sample with range w in $q = w/s$, and (ii) ν degrees of freedom for the denominator mean square, which are also those of the independent value of s in $q = w/s$.

In fact, for any given n, the ratio $F(n-1, \nu)/q^2$ is almost constant for all values of ν and could be easily obtained by interpolation for the intermediate values of $\nu = 5$ and 7 and converted to q with the help of the published tables of $F(n-1, \nu)$ (*Biometrika* (1943), **33**, 73). Further, this property made it possible to apply an additional check to all values of q in the tables. Only for the lines $\nu = 5, 7$ were discrepancies found exceeding 4 units in the last figure, and the corrected values for these lines only are therefore given above. Small end-figure adjustments of other values will, however, be made when the full corrected table is published as part of the forthcoming *Biometrika Tables for Statisticians*.

H. O. HARTLEY

(2) J. DRAPER, *Biometrika* (1952), **39**, 299.

In the last column of the table half-way down the page, *for* 2·753 *read* 2·750.

---

* We are indebted to Mr George W. Thomson of the Ethyl Corporation, Michigan, for drawing our attention to the discrepancies between the values of q and $\sqrt{2}t$ for $\nu = 5$ and 7.

VOLUME 40, PARTS 3 AND 4 DECEMBER 1953

THE POPULATION FREQUENCIES OF SPECIES AND THE ESTIMATION OF POPULATION PARAMETERS

BY I. J. GOOD

A random sample is drawn from a population of animals of various species. (The theory may also be applied to studies of literary vocabulary, for example.) If a particular species is represented r times in the sample of size N, then r/N is not a good estimate of the population frequency, p, when r is small. Methods are given for estimating p, assuming virtually nothing about the underlying population. The estimates are expressed in terms of smoothed values of the numbers n_r ($r=1, 2, 3, \ldots$), where n_r is the number of distinct species that are each represented r times in the sample. (n_r may be described as 'the frequency of the frequency r'.) Turing is acknowledged for the most interesting formula in this part of the work. An estimate of the proportion of the population represented by the species occurring in the sample is an immediate corollary. Estimates are made of measures of heterogeneity of the population, including Yule's 'characteristic' and Shannon's 'entropy'. Methods are then discussed that do depend on assumptions about the underlying population. It is here that most work has been done by other writers. It is pointed out that a hypothesis can give a good fit to the numbers n_r but can give quite the wrong value for Yule's characteristic. An example of this is Fisher's fit to some data of Williams's on Macrolepidoptera.

1. *Introduction.* We imagine a random sample to be drawn from an infinite population of animals of various species. Let the sample size be N and let n_r distinct species be each represented exactly r times in the sample, so that

$$\sum_{r=1}^{\infty} rn_r = N. \tag{1}$$

The sample tells us the values of $n_1, n_2, \ldots$, but not of n_0. In fact it is not quite essential that n_0 should be finite though we shall find it convenient to suppose that it is.

We shall suggest a method of estimating, among other things,

(i) the population frequency of each species;

(ii) the total population frequency of all species represented in the sample, or, as we may say, 'the proportion of the population represented by (the species occurring in) the sample';

(iii) various general population parameters measuring heterogeneity, including 'entropy'. By 'general' parameters we mean parameters defined without reference to any special form of hypothesis. In § 7 we shall consider the estimation of parameters for hypotheses of special forms.

Our results are applicable, for example, to studies of literary vocabulary, of accident proneness and of chess openings, but for definiteness we formulate the theory in terms of species of animals.

The formula (2) was first suggested to me, together with an intuitive demonstration, by Dr A. M. Turing several years ago. Hence a very large part of the credit for the present paper should be given to him, and I am most grateful to him for allowing me to publish this work.

Reasonably precise conditions under which our general results are applicable will be given in § 4, but we state at once that the larger is n_1 the more applicable the results. When n_1 is large, n_0 will also be large, but we shall not for the most part attempt to estimate it. There will be a fleeting reference to the estimation of n_0 at the end of § 5 and a few more references in §§ 7 and 8. (See, for example, equation (73).) For populations of known finite size, the problem has been considered by Goodman (1949). He proved that if the sample size is not less than the maximum number of individuals in the population belonging to a single species, then there is only one unbiased estimate of n_0 and he found it. He also pointed out that the unbiased estimate is liable to be unreasonable and suggested some alternative estimates that are always reasonable. There is practically no overlapping between the present work and that of Goodman.

Jeffreys (1948, §3·23) has discussed what is superficially the same problem as (i) above, under the heading 'multiple sampling'. He refers to some earlier work of Johnson (1932). The methods of Johnson and Jeffreys depend on assumptions that, as Jeffreys himself points out, are not always acceptable. Moreover, their methods are not intended to be applicable when n_0 is unknown. The matter is taken up again in §2.

Other work on the frequencies of species has been mainly concerned with the fitting of particular distributions to the data, with or without a theoretical explanation of why these distributions might be expected to be suitable. See, for example, Anscombe (1950), Chambers & Yule (1942), Corbet, Fisher & Williams (1943), Greenwood & Yule (1920), Newbold (1927), Preston (1948), Yule (1944) and Zipf (1932). The methods of the first six sections of the present paper are largely independent of the distributions of population frequencies.

We shall be largely concerned with q_r, the population frequency of an arbitrary species that is represented r times in the sample. We shall use the notation $\mathscr{E}(q_r)$ for the expected value of q_r, in a sense to be explained in §2. Our main result, expressed rather loosely, is that the expected value of q_r is r^*/N, where

$$r^* \simeq (r+1)\, n_{r+1}/n_r. \tag{2}$$

(The symbol '$\simeq$' is used throughout to mean 'is approximately equal to'.) More precisely the n_r's should first be smoothed before applying formula (2). Smoothing is briefly discussed in §3 with examples in §8. If the smoothed values are denoted by $n_1', n_2', n_3', \ldots$, then the more accurate form of equation (2) is

$$r^* \simeq (r+1)\, n'_{r+1}/n'_r. \tag{2'}$$

The reader will find it instructive to consider the special case when n'_r is of the Poisson form $s\,e^{-a}a^r/r!$ Then r^* reduces to a constant.

The formula (2) can be generalized to give higher moments of q_r. In fact

$$\mathscr{E}(q_r^m) \simeq \frac{(r+m)^{(m)}}{N^m}\frac{n_{r+m}}{n_r} \quad (r = 1, 2, 3, \ldots;\ m = 0, 1, 2, \ldots), \tag{3}$$

where $t^{(m)} = t(t-1)\ldots(t-m+1)$. We can also write (3) in the form

$$\mathscr{E}(q_r^m) \simeq \mathscr{E}(q_r)\,\mathscr{E}(q_{r+1}) \ldots \mathscr{E}(q_{r+m-1}). \tag{4}$$

Moreover, the variance of q_r is

$$\begin{aligned} V(q_r) &\simeq \frac{(r+1)(r+2)}{N^2}\frac{n_{r+2}}{n_r} - \left(\frac{r+1}{N}\frac{n_{r+1}}{n_r}\right)^2 \\ &\simeq \mathscr{E}(q_r)\,[\mathscr{E}(q_{r+1}) - \mathscr{E}(q_r)]. \end{aligned} \tag{5}$$

An immediate deduction from (2) is that *the expected total chance of all species that are each represented r times* ($r \geqslant 1$) *in the sample is approximately*

$$(r+1)\, n_{r+1}/N. \tag{6}$$

Hence also the expected total chance of all species that are represented r times or more in the sample is approximately

$$N^{-1}\{(r+1)\,n_{r+1}+(r+2)\,n_{r+2}+\ldots\}. \tag{7}$$

In particular, the expected total chance of all species represented at all in the sample is approximately

$$N^{-1}(2n_2+3n_3+\ldots) = 1-n_1/N. \tag{8}$$

We may say that the proportion of the population represented by the sample is approximately $1-n_1/N$, and *the chance that the next animal sampled will belong to a new species is approximately*

$$n_1/N. \tag{9}$$

(Thus (6) is true even if $r = 0$.)

The results (6), (7), (8) and (9) are improved in accuracy by writing the respective formulae as

$$\frac{(r+1)\,n'_{r+1}n_r}{n'_rN}, \tag{6$'$}$$

$$N^{-1}\left\{\frac{(r+1)\,n'_{r+1}n_r}{n'_r}+\frac{(r+2)\,n'_{r+2}n_{r+1}}{n'_{r+1}}+\ldots\right\}, \tag{7$'$}$$

$$N^{-1}\left\{\frac{2n'_2n_1}{n'_1}+\frac{3n'_3n_2}{n'_2}+\ldots\right\} \tag{8$'$}$$

and

$$\frac{n'_1n_0}{n'_0N}. \tag{9$'$}$$

In most applications this last expression will be extremely close to n'_1/N, and this in its turn will often be very close to n_1/N. It follows that (8′) and (9′) are practically the same as (8) and (9). For the sake of mathematical consistency, the smoothing should be such that (8′) and (9′) add up to 1.

An index of notations used in a fixed sense is given in § 9.

I am grateful, and my readers will also be grateful, to Prof. M. G. Kendall for forcing me to clarify some obscurities, especially in §§ 1 and 2.

2. *Proofs*. Let the number of species in the population be s, which we suppose is finite. This is the same supposition as that n_0 is finite. Our results as far as § 6 would be practically unchanged if s were enumerably infinite, but the proofs are more rigorous when it is finite. Let the population frequencies of the species be, in some order, $p_1, p_2, \ldots, p_s$, where

$$p_1+p_2+\ldots+p_s = 1, \quad n_0+n_1+\ldots = s.$$

Let H, or more explicitly $H(p_1, p_2, \ldots, p_s)$, be the statistical hypothesis asserting that $p_1, p_2, \ldots, p_s$ are the population frequencies. We shall discuss the expectation of n_r, given H. It may be objected that the expectation of n_r is simply the observed number n_r, whatever the information, and this objection would be logically correct. Strictly we should introduce extra notation, say $\nu_{r,N}$, for the random variable that is the frequency of the frequency r in a random sample of size N. Then we could introduce the notation $\mathscr{E}(\nu_{r,N}\mid H)$ for the expectation of $\nu_{r,N}$ given H. (Logically this expectation would remain unaffected if particular values of $n_1, n_2, n_3, \ldots$ were given.) In order to avoid the extra notation $\nu_{r,N}$ we shall write $\mathscr{E}(n_r)$ or $\mathscr{E}(n_r\mid H)$ or $\mathscr{E}_N(n_r\mid H)$ instead of $\mathscr{E}(\nu_{r,N}\mid H)$. Confusion can be avoided by reading $\mathscr{E}_N(n_r\mid H)$ as 'the expectation of the frequency of the frequency r when H is given and when the sample size is N'. Similarly, we write $V(n_r) = V(n_r \mid H) = V_N(n_r \mid H)$ for the variance of $\nu_{r,N}$ given H and $\mathscr{E}_N(n_r^2 \mid H)$, etc., for $\mathscr{E}(\nu_{r,N}^2 \mid H)$.

We recall the theorem that an expectation of a sum is the sum of the expectations. It follows that $\mathscr{E}_N(n_r \mid H)$ is the sum over all s species of the probabilities that each will occur r times, given H. So

$$\mathscr{E}_N(n_r \mid H) = \mathscr{E}(n_r \mid H) = \mathscr{E}(n_r)$$

$$= \sum_{\mu=1}^{s} \binom{N}{r} p_\mu^r (1-p_\mu)^{N-r}. \tag{10}$$

In particular

$$\mathscr{E}_N(n_0 \mid H) = \sum_{\mu=1}^{s} (1-p_\mu)^N. \tag{11}$$

If s were infinite this series would diverge. The divergence would be appropriate since n_0 would also be infinite.

Now suppose that in a sample of size N a particular species occurs r times ($r = 0, 1, 2, \ldots$). We shall consider the final (*posterior*) probability that this species is the μth one (of population frequency p_μ). For the sake of rigour it is necessary to define more precisely how the species is selected for consideration. We shall suppose that it is sampled 'at random', or rather equiprobably, from the s species, and that then its number of occurrences in the sample is counted. Thus the initial (*prior*) probability that the species is the μth one is $1/s$. If the species is the μth one then the likelihood that the observed number of occurrences is r is

$$\binom{N}{r} p_\mu^r (1-p_\mu)^{N-r}.$$

We write q_r for the (unknown) population frequency of an arbitrary species that is represented r times in the sample. The final probability that the species is the μth one can be written as $P(q_r = p_\mu \mid H)$ provided that the p_μ's are unequal. (If any of the p_μ's are equal they can be adjusted microscopically so as to be made unequal. These adjustments will have no practical effect.) We may at once deduce the final probability that the species is the μth one by using Bayes's theorem in the form that the final probabilities are proportional to the initial ones times the likelihoods. We find that

$$P\,(q_r = p_\mu \mid H) = \frac{p_\mu^r (1-p_\mu)^{N-r}}{\sum\limits_{\mu=1}^{s} p_\mu^r (1-p_\mu)^{N-r}}. \tag{12}$$

It follows that for any positive integer m,

$$\mathscr{E}(q_r^m \mid H) = \frac{\sum\limits_{\mu=1}^{s} p_\mu^{r+m} (1-p_\mu)^{N-r}}{\sum\limits_{\mu=1}^{s} p_\mu^r (1-p_\mu)^{N-r}} \tag{13}$$

$$= \frac{(r+m)^{(m)}}{(N+m)^{(m)}} \frac{\mathscr{E}_{N+m}(n_{r+m} \mid H)}{\mathscr{E}_N(n_r \mid H)}, \tag{14}$$

in view of (10) and of (10) with N replaced by $N+m$. Immediate consequences of (14) are the basic result

$$\mathscr{E}(q_r \mid H) = \frac{r+1}{N+1} \frac{\mathscr{E}_{N+1}(n_{r+1} \mid H)}{\mathscr{E}_N(n_r \mid H)} \quad (r = 0, 1, 2, \ldots) \tag{15}$$

and

$$V(q_r \mid H) = \frac{\mu'_{r,2,N}\mu'_{r,0,N} - \mu'^2_{r,1,N}}{\mu'^2_{r,0,N}}, \tag{16}$$

where
$$\mu'_{r,t,N} = \frac{(r+t)!}{(N+t)!}\,\mathscr{E}_{N+t}(n_{r+t} \mid H) \tag{17}$$

$$= \frac{1}{(N-r)!}\sum_{\mu=1}^{s} p_\mu^{r+t}(1-p_\mu)^{N-r} = \frac{r!}{N!}\,\mathscr{E}_N(n_r \mid H)\,\mathscr{E}(q_r^t \mid H) \tag{18}$$

by (10) and (14). It is clear from either form of (18) that the numbers $\mu'_{r,t,N}$ $(t = 0, 1, 2, \ldots)$ form a sequence of moment constants and therefore satisfy Liapounoff's inequality. (See, for example, Good (1950*a*), or Uspensky (1937).) This checks that the right side of (16) is positive, as it should be being a variance. [It is obvious incidentally that (16) would be true with $\mu'_{r,t,N}$ defined as $\mathscr{E}(q_r^t \mid H)$ times any expression independent of t.]

We can now approximate the formulae (14) and (15) by replacing $\mathscr{E}_{N+m}(n_{r+m} \mid H)$ by the observed value, n_{r+m}, in the sample of size N, or rather by the smoothed value n'_{r+m}. If m is very small compared with N, if n_r and n_{r+m} are not too small and if the sequence $n_1, n_2, n_3, \ldots$ is smoothed in the neighbourhood of n_r and n_{r+m}, then we may expect the approximations to be good. We thus obtain all the approximate results of § 1. Note that when the approximation is made of replacing $\mathscr{E}_{N+m}(n_{r+m} \mid H)$ by n'_{r+m} we naturally also change the notation $\mathscr{E}(q_r^m \mid H)$ to $\mathscr{E}(q_r^m)$. For the results become roughly independent of H unless the n_r's are too small to smooth. Observe that $\mathscr{E}(q_r^m \mid H)$ does not depend on the sample, unless H is itself determined by using the sample. On the other hand, $\mathscr{E}(q_r^m)$ does depend on the sample. This may seem a little paradoxical and the following explanation is perhaps worth giving. When we select a particular sequence of smoothed values $n'_1, n'_2, n'_3, \ldots$ we are virtually accepting a particular hypothesis H, say $H\{N;\, n'_1, n'_2, n'_3, \ldots\}$, with curly brackets. (I do not think that this hypothesis is usually a simple statistical hypothesis.) Then $\mathscr{E}(q_r^m)$ can be regarded as a shorthand for $\mathscr{E}(q_r^m \mid H\{N;\, n'_1, n'_2, n'_3, \ldots\})$. (If $H\{\ldots\}$ is not a simple statistical hypothesis this last expression could in theory be given a definite value by assuming a definite distribution of probabilities of the simple statistical hypotheses of which H is a disjunction.) When we regard the smoothing as reasonably reliable we are virtually taking $H\{N;\, n'_1, n'_2, n'_3, \ldots\}$ for granted, as an approximation, so that it can be omitted from the notation without serious risk of confusion. In order to remind ourselves that there is a logical question that is obscured by the notation, we may describe $\mathscr{E}(q_r^m)$ as say a 'credential expectation'.

If a specific H is accepted it is clearly not necessary to use the approximations since equation (13) can then be used directly. Similarly, if H is assumed to be selected from a superpopulation, with an assigned probability density, then again it is theoretically possible to dispense with the approximations. In fact if the 'point' $(p_1, p_2, \ldots, p_s)$ is assumed to be selected from the 'simplex' $p_1 + p_2 + \ldots + p_s = 1$, with probability density proportional to $(p_1 p_2 \ldots p_s)^{k-1}$, where k is a constant, then it is possible to deduce Johnson's estimate $q_r = (r+k)/(N+ks)$. Jeffreys's estimate is the special case $k = 1$, when the probability density is uniform. Jeffreys suggests conditions for the applicability of his estimate, but these conditions are not valid for our problem in general. This is clear if only because we do not assume s to be known.

Jeffreys assumes explicitly that all ordered partitions of N into s non-negative parts are initially equally probable, while Johnson assumes that the probability that the next individual sampled will be of a particular species depends only on N and on the number of times that that species is already represented in the sample. Clearly both methods ignore any information that can be obtained from the entire set of frequencies of all species.

The ignored information is considerable when it is reasonable to smooth the frequencies of the frequencies.

3. *Smoothing.* The purpose of smoothing the sequence $n_1, n_2, n_3, \ldots$ and replacing it by a new sequence $n'_1, n'_2, n'_3, \ldots$, is to be able to make sensible use of the exact results (14) and (15). Ignoring the discrepancy between $\mathscr{E}_N$ and $\mathscr{E}_{N+m}$, the best value of n'_r would be $\mathscr{E}_N(n_r \mid H)$, where H is true. One method of smoothing would be to assume that $H = H(p_1, p_2, \ldots, p_s)$ belongs to some particular set of possible H's, to determine one of these, say H_0, by maximum likelihood and then to calculate n'_r as $\mathscr{E}_N(n_r \mid H_0)$. This method is closely related to that of Fisher in Corbet *et al.* (1943). Since one of our aims is to suggest methods which are virtually distribution-free, it would be most satisfactory to carry out the above method using all possible H's as the set from which to determine H_0. Unfortunately, this theoretically satisfying method leads to a mathematical problem that I have not solved.

It is worth noticing that the sequence $\{\mathscr{E}_N(n_r \mid H)\}$ $(r = 0, 1, 2, \ldots)$ has some properties invariant with respect to changes in H. Ideally the sequence $\{n'_r\}$ should be forced to have these invariant properties. In particular the sequence $\{\mu'_{r,t,N}\}$ $(t = 0, 1, 2, \ldots)$, defined by (17), is a sequence of moment constants. But if $t = o(\sqrt{N})$, then $N^{-t}(r+t)!\, n'_{r+t} \simeq \mu'_{r,t,N}$, so that if $t = o(\sqrt{N})$ we can assume that the sequence $r!\, n'_r$ is a sequence of moment constants and satisfies Liapounoff's inequalities. But this simply implies that $0^*, 1^*, 2^*, \ldots, t^*$ forms an increasing sequence (see equation (2′)), a result which is intuitively obvious even without the restriction $t = o(\sqrt{N})$. (Indeed, the argument could be reversed in order to obtain a new proof of Liapounoff's inequality.) We also intuitively require that $0^*, 1^*, 2^* \ldots$ should itself be a 'smooth' sequence.

Since the sequence $\{\mu'_{r,t,N}\}$ $(t = 0, 1, 2, \ldots)$ is a sequence of moment constants of a probability distribution it follows from Hardy (1949, §11·8) that the sequence is 'totally increasing', i.e. that all its finite differences are non-negative. This result is unfortunately too weak to be useful for our purposes, but it may be possible to make use of some other theorems concerning moment constants. This line of research will not be pursued in the present paper.

A natural principle to adopt when smoothing is that

$$\chi^2 = \sum_{t=1}^{r} \frac{(n'_t - n_t)^2}{V(n_t \mid H)} \tag{19}$$

should not be significant with r degrees of freedom. In §5 we shall obtain an approximate formula for $V(n_r \mid H)$, applicable when $r^2 = o(N)$. The chi-squared test will therefore be applicable when $r^2 = o(N)$. [See formulae (22), (25), (26) and, for particular H's, (65), (85), (86).]

Another similar principle can be understood by thinking of the histogram of n_r as several piles of pennies, n_r pennies in the rth pile. We may visualize the smoothing as the moving of pennies from pile to pile, and we may reasonably insist that pennies moved to the rth pile should not have been moved much further horizontally than a distance $\sqrt{r}$ and almost never further than $2\sqrt{r}$. For $r = 0$ we would not insist on this rule, i.e. we do not insist that $\sum_{r=1}^{\infty} n'_r = \sum_{r=1}^{\infty} n_r$. The analogy with piles of pennies amounts to saying that a species that 'should' have occurred r times is unlikely to have occurred less than $r - \sqrt{r}$ or more than $r + \sqrt{r}$ times.

Let $N' = \Sigma r n_r'$. It seems unnecessary to insist on $N' = N$, provided that N is replaced by N' in such formulae as $\mathscr{E}(q_r) \simeq r^*/N$. It will be convenient, however, in § 6 to assume $N' = N$.

For some applications very little smoothing will be required, while for others it may be necessary to use quite elaborate methods. For example, we could

(i) Smooth the n_r's for the range of values of r that interests us, holding in mind the above chi-squared test and the rule concerning $\sqrt{r}$. The smoothing techniques may include the use of freehand curves. Rather than working directly with $n_1, n_2, n_3, \ldots$ it may be found more suitable to work with the cumulative sums $n_1, n_1+n_2, n_1+n_2+n_3, \ldots$ or with the cumulative sums of the rn_r or with the logarithms $\log n_1$, $\log n_2$, $\log n_3, \ldots$. There is much to be said for working with the numbers $\sqrt{n_1}, \sqrt{n_2}, \sqrt{n_3}, \ldots$. For if we assume that $V(n_r \mid H)$ is approximately equal to n_r (and in view of (26) and (27) of § 3 this approximation is not on the whole too bad), then it would follow that the standard deviation of $\sqrt{n_r}$ is of the order of $\frac{1}{2}$ and therefore largely independent of r. Hence graphical and other smoothing methods can be carried out without having constantly to hold in mind that $|n_r' - n_r|$ can reasonably take much larger values when n_r is large than when it is small. [The square-root transformation for a Poisson variable, x, was suggested by Bartlett (1936) in order to facilitate the analysis of variance. He showed also that the transformation $\sqrt{(x+\frac{1}{2})}$ leads to an even more constant variance. Anscombe (1948) proved that $\sqrt{(x+\frac{3}{8})}$ has the most nearly constant variance of any variable of the form $\sqrt{(x+c)}$, namely, $\frac{1}{4}$, when the mean of x is large. He attributes this result to A. H. L. Johnson.]

(ii) Calculate $(r+1)\, n_{r+1}'/n_r'$.

(iii) Smooth these values getting, say, r^*.

(iv) Possibly use the values of r^* to improve the smoothing of the n_r's. If this makes a serious difference it will be necessary to check again that the chi-squared test and the $\sqrt{r}$ rule have not been violated.

(v) Light can be shed on the reliability of the estimates of the q_r's, etc., if the data are smoothed two or three times, possibly by different people.

In short, the estimation of the q_r's should be done in such a way as to be consistent with the axioms of probability and also with any intuitive judgements that the users of the method are not prepared to abandon or to modify. (This recommendation applies to much more general theoretical scientific work, though there are rare occasions when it may be preferred to abandon the axioms of a science.)

An objection could be raised to the methods of smoothing suggested in the present section. It could be argued that all smoothing methods indirectly assume something about the distribution p_μ, and that one might just as well apply the method of Greenwood & Yule (1920) and its modification by Corbet *et al.* (1943) of assuming a distribution of Pearson's Type III, $Ap^\alpha e^{-\beta p}$, or of some other form. Our reply would be that smoothing can be done by making only *local* assumptions, for example, that the square root of $\mathscr{E}(n_r \mid H)$, as a function of r, is approximately 'parabolic' for any nine consecutive values of r. Moreover, it may often be more convenient to apply the general methods of the present section than to attempt to find an adequate hypothesis, H.

4. *Conditions for the applicability of the results of* §§ 1 *and* 2. The condition for the applicability of the results of §§ 1 and 2 is that the user of the methods should be satisfied with his approximations to $\mathscr{E}_{N+m}(n_{r+m} \mid H)$ corresponding to the values of r and m used in the application. This condition is clearly correct, since equation (14) is exact. In particular, if

n_1 is large enough the user would be quite happy to deduce (9) from (15) with $r = 0$. Similarly, he will be satisfied with the estimates of say q_1, q_2 and q_3 provided he is satisfied with the smoothed values (n'_1, n'_2, n'_3, n'_4) of n_1, n_2, n_3 and n_4.

5. *The variance of* n_r. For the application of the chi-squared test described in § 3 we need to know more about $V(n_r)$. We begin by obtaining an exact formula for $V(n_r \mid H) = V_N(n_r \mid H)$ and we then make approximations that justify the omission of the symbol H from the notation. It is convenient to introduce the random variable $x_{\mu,r} = x_\mu$ that is defined as 1 if the 'μth species' (of population frequency p_μ) occurs precisely r times in a sample of size N (H being given), otherwise $x_\mu = 0$. Clearly $P\,(x_\mu = 1 \mid H) = \binom{N}{r} p_\mu^r (1-p_\mu)^{N-r}$. Now

$$\begin{aligned}
\mathscr{E}(n_r^2 \mid H) &= \mathscr{E}(\sum_\mu x_\mu)^2 \\
&= \sum_{\mu,\nu} \mathscr{E}(x_\mu x_\nu) \\
&= \sum_\mu \mathscr{E}(x_\mu) + \sum_{\mu,\nu}^{\mu\neq\nu} \mathscr{E}(x_\mu x_\nu) \\
&= \mathscr{E}(n_r \mid H) + \frac{N!}{r!\,r!\,(N-2r)!} \sum_{\mu,\nu}^{\mu\neq\nu} p_\mu^r p_\nu^r (1-p_\mu-p_\nu)^{N-2r}.
\end{aligned} \tag{20}$$

This is exact. We now make some approximations of the sort used in deriving the Poisson distribution from the binomial. We get, assuming r^2/N, rp_μ and rp_ν to be small,

$$\binom{N}{r} p_\mu^r (1-p_\mu)^{N-r} \simeq \frac{(Np_\mu)^r\, e^{-Np_\mu}}{r!} = a_\mu, \text{ say,}$$

and

$$\frac{N!}{r!\,r!\,(N-2r)!} p_\mu^r p_\nu^r (1-p_\mu-p_\nu)^{N-2r} \simeq a_\mu a_\nu.$$

Moreover, it is intuitively clear that terms for which p_μ or p_ν is far from r/N can make no serious contribution to the summation in (20). Hence, if $r^2 = o(N)$,

$$\begin{aligned}
\mathscr{E}(n_r^2 \mid H) &\simeq \mathscr{E}(n_r \mid H) + \sum_{\mu,\nu}^{\mu\neq\nu} a_\mu a_\nu \\
&\simeq \mathscr{E}(n_r \mid H) + \{\mathscr{E}(n_r \mid H)\}^2 - \sum_\mu a_\mu^2.
\end{aligned}$$

Therefore the variance of n_r for samples of size N is

$$V_N(n_r \mid H) \simeq \mathscr{E}_N(n_r \mid H) - \sum_\mu \frac{(Np_\mu)^{2r}\, e^{-2Np_\mu}}{(r!)^2} \tag{21}$$

$$\simeq \mathscr{E}_N(n_r \mid H) - \frac{1}{2^{2r}} \binom{2r}{r} \mathscr{E}_{2N}(n_{2r} \mid H). \tag{22}$$

Formulae (21) and (22) are elegant but need further transformation, when H is unknown, before they can be used for calculation. Notice first that there are n_u species whose expected population frequencies are q_u $(u = 0, 1, 2, \ldots)$. Hence we have for $r = 0, 1, 2, \ldots;\ r^2 = o(N)$,

$$\begin{aligned}
V(n_r \mid H) &\simeq \mathscr{E}(n_r \mid H) - \sum_{u=0}^{\infty} n_u \left\{\frac{(Nq_u)^r\, e^{-Nq_u}}{r!}\right\}^2 \\
&= \mathscr{E}(n_r \mid H) - \sum_{u=0}^{\infty} n_u \left\{\frac{u^{*r}\, e^{-u^*}}{r!}\right\}^2.
\end{aligned} \tag{23}$$

Similarly and rather more simply, when $r^2 = o(N)$,

$$\mathscr{E}(n_r \mid H) \simeq \frac{1}{r!} \sum_{u=0}^{\infty} n_u u^{*r} e^{-u^*}. \tag{24}$$

Now for any positive x, $x^r e^{-x} \leqslant r^r e^{-r}$, so

$$\mathscr{E}(n_r \mid H) \mathrel{\overset{\frown}{\geq}} V(n_r \mid H) \mathrel{\overset{\frown}{\geq}} \mathscr{E}(n_r \mid H) \left(1 - \frac{r^r e^{-r}}{r!}\right). \tag{25}$$

Using Stirling's formula for $r-1$ we have

$$\mathscr{E}(n_r \mid H) \mathrel{\overset{\frown}{\geq}} V(n_r \mid H) \mathrel{\overset{\frown}{\geq}} \mathscr{E}(n_r \mid H) \left(1 - \frac{1}{\sqrt{(2\pi r)}}\right) \quad (r = 2, 3, \ldots), \tag{26}$$

while

$$\mathscr{E}(n_1 \mid H) \mathrel{\overset{\frown}{\geq}} V(n_1 \mid H) \mathrel{\overset{\frown}{\geq}} \mathscr{E}(n_1 \mid H)(1 - e^{-1}) \tag{27}$$

(see also formula (65) in §7). Now the most desirable value for n'_r would be $\mathscr{E}(n_r \mid H)$ where H is true, so if our smoothing of the n_r's is to be satisfactory for any particular values of r small compared with $\sqrt{N}$ we may write

$$n'_r \simeq \sum_{u=0}^{\infty} n_u \frac{u^{*r} e^{-u^*}}{r!}, \tag{28}$$

and these approximate equations may be used as a test of consistency for the values of n'_r and u^*. Indeed, it may be possible iteratively to solve equations (28) combined with (2′) and thus very systematically to obtain estimates of n'_r and r^* for values of r small compared with $\sqrt{N}$. This iterative process may possibly lead to estimates of n'_0 and 0^*, but I have not yet tried out the process. For most applications the less systematic methods previously described will probably prove to be adequate, and any smoothing obtained by these methods can be partially tested by means of χ^2 in the form (19), together with the inequalities (26) and (27). (See also the remarks following equations (65) and (87).)

6. *Estimation of some population parameters, including entropy.* Let us consider the population parameters

$$c_{m,n} = \sum_{\mu=1}^{s} p_\mu^m (-\log p_\mu)^n \quad (m, n = 0, 1, 2, \ldots), \tag{29}$$

which can be regarded as measures of heterogeneity of the population. The sequence $c_{0,0} = 1$, $c_{1,0} = s$, $c_{2,0}$, $c_{3,0}$, … may be called the 'moment constants' of the population, while $c_{1\,1}$ is called the 'entropy' in the modern theory of communication (see Shannon, 1948). More generally, $c_{1,n}$ is the moment about zero of the amount of information from each selection of an animal (or word), where 'amount of information' is here used in the sense of Good (1950*b*, p. 75), i.e. as minus the logarithm of a probability. (The last sentence of p. 75 of this reference is incorrect, as Prof. M. S. Bartlett has pointed out.) We find it no more difficult to give estimates of $c_{m,n}$ than of $c_{1,n}$, at any rate when $n = 0$ or 1.

It is an immediate consequence of (10) that an unbiased estimate of $c_{m,0}$ is

$$\hat{c}_{m,0} = \frac{1}{N^{(m)}} \sum_r r^{(m)} n_r. \tag{30}$$

$\hat{c}_{2,0}$ is in effect used by Yule (1944) to measure the heterogeneity of samples of vocabulary, and he calls $10{,}000\hat{c}_{2,0}(1 - 1/N)$ the 'characteristic' of the material. The sequence of all sampling moments of $\hat{c}_{2,0}$ involves all the population parameters $c_{m,0}$. For example, as pointed out by Simpson (1949), for large N,

$$V(\hat{c}_{2,0}) \simeq \frac{4}{N}(c_{3,0} - c_{2,0}^2). \tag{30 A}$$

Unbiased statistics are rather unfashionable nowadays, partly because they can take impossible values. For example, $\hat{c}_{m,0}$ could vanish, although it is easy to see that $c_{m,0} \geqslant s^{-(m-1)}$. (Compare Good (1950*b*, p. 103), where estimates of $c_{m,0}$ are implicit for general multinomial distributions, no attempt being made to smooth the n_r's.) We shall find estimates of $c_{m,1}$ and also estimates of $c_{m,0}$ that are at least sometimes better than $\hat{c}_{m,0}$.

We have

$$c_{m,0} = \frac{1}{N^{(m)}} \sum_r r^{(m)} \mathscr{E}(n_r \mid H), \tag{31}$$

since this is in effect what is meant by saying that $\hat{c}_{m,0}$ is an unbiased estimate of $c_{m,0}$. If the statistician is satisfied with his smoothing, i.e. if he assumes that $n'_r \simeq \mathscr{E}(n_r \mid H)$, and if he has forced $N' = N$, then he can estimate $c_{m,0}$ as

$$\tilde{c}_{m,0} = \frac{1}{N^{(m)}} \sum_r r^{(m)} n'_r, \tag{32}$$

and he will be prepared to assume that this is a more efficient estimate than $\hat{c}_{m,0}$. More generally if the smoothing is satisfactory for $r = 1, 2, \ldots, t$ but not for all larger values of r, then a good estimate of $c_{m,0}$ will be $\tilde{c}_{m,0}(t)$, where

$$\tilde{c}_{m,0}(t) = \frac{1}{N^{(m)}} \left\{ \sum_{r=m}^{t} r^{(m)} n'_r + \sum_{r=t+1}^{\infty} r^{(m)} n_r \right\}. \tag{33}$$

We shall next consider estimates $\tilde{c}_{m,1}$ of $c_{m,1}$. We shall begin by proving that (exactly)

$$c_{m,1} = \frac{1}{N^{(m)}} \sum_r r^{(m)} \mathscr{E}(n_r \mid H) \left\{ \frac{1}{r+1} + \frac{1}{r+2} + \ldots + \frac{1}{N-r} - \frac{d}{dr} \log \mathscr{E}(n_r \mid H) - \mathscr{E}[\log(1-q_r) \mid H] \right\}. \tag{34}$$

The differential coefficient in this expression is made meaningful by means of a suitable definition of $\mathscr{E}(n_r \mid H)$ for non-integral values of r. This definition is obtained from equation (10) by writing $\Gamma(N+1)/\Gamma(r+1)\,\Gamma(N-r+1)$ instead of $\binom{N}{r}$.

In order to prove (34) we shall need the following generalization of (13), valid for any function $f(.)$:

$$\mathscr{E}(n_r \mid H)\, \mathscr{E}[f(q_r) \mid H] = \binom{N}{r} \sum_\mu p_\mu^r (1-p_\mu)^{N-r} f(p_\mu). \tag{35}$$

We also require the following property of the gamma function. If b is a non-negative integer,

$$\frac{\Gamma'(b+1)}{\Gamma(b+1)} = 1 + \frac{1}{2} + \ldots + \frac{1}{b} - \gamma, \tag{36}$$

where $\gamma = 0{\cdot}577215\ldots$ is the Euler-Mascheroni constant. (See, for example, Jeffreys & Jeffreys (1946, §15·04).) It follows from (10) and (36) that

$$\begin{aligned} \frac{d}{dr} \mathscr{E}(n_r \mid H) &= \binom{N}{r} \sum_\mu p_\mu^r (1-p_\mu)^{N-r} \left(\frac{1}{r+1} + \frac{1}{r+2} + \ldots + \frac{1}{N-r} + \log \frac{p_\mu}{1-p_\mu} \right) \\ &= \mathscr{E}(n_r \mid H) \left\{ \frac{1}{r+1} + \ldots + \frac{1}{N-r} + \mathscr{E}\left(\log \frac{q_r}{1-q_r} \mid H \right) \right\}, \end{aligned}$$

by (35). Therefore

$$\mathscr{E}(n_r\mid H)\left(\frac{1}{r+1}+\dots+\frac{1}{N-r}\right)-\frac{d}{dr}\mathscr{E}(n_r\mid H)-\mathscr{E}(n_r\mid H)\,\mathscr{E}[\log(1-q_r)\mid H]$$
$$=\mathscr{E}(n_r\mid H)\,\mathscr{E}(-\log q_r\mid H)=\binom{N}{r}\sum_{\mu}p_\mu^r(1-p_\mu)^{N-r}(-\log p_\mu),$$

by (35) again. Multiplying by $r^{(m)}/N^{(m)}$ and summing with respect to r, we find that the right-hand side of (34) equals

$$\sum_{\mu}\sum_{r}\binom{N-m}{r-m}p_\mu^{r-m}(1-p_\mu)^{N-m-(r-m)}p_\mu^{m}(-\log p_\mu)=c_{m,1}$$

as asserted.

$c_{m,2}$ can be evaluated in a similar manner by first writing down $\left(\frac{d}{dr}\right)^2\mathscr{E}(n_r\mid H)$, but the result is complicated and will be omitted.

As in the estimation of $c_{m,0}$, if the statistician is satisfied with his smoothing, then he can write

$$c_{m,1}\bumpeq\frac{1}{N^{(m)}}\sum_{r}r^{(m)}n'_r\left\{\frac{1}{r+1}+\frac{1}{r+2}+\dots+\frac{1}{N-r}-\frac{d}{dr}\log n'_r-\mathscr{E}[\log(1-q_r)\mid H]\right\}.$$

If N is large the approximation can be written

$$c_{m,1}\bumpeq\frac{1}{N^{(m)}}\sum_{r}r^{(m)}n'_r\left\{\log N-\left(1+\frac{1}{2}+\dots+\frac{1}{r}-\gamma\right)-\frac{d}{dr}\log n'_r-\mathscr{E}\left(\log\frac{1-q_r}{1-\frac{r}{N}}\mid H\right)\right\}.$$

Now it is intuitively clear that $\mathscr{E}\left(q_r-\frac{r}{N}\right)$, which equals $\frac{r^*-r}{N}$, must be $O\left(\frac{\sqrt{r}}{N}\right)=o(N)$, and therefore

$$c_{m,1}\bumpeq\frac{1}{N^{(m)}}\sum_{r}r^{(m)}n'_r\left\{\log N-\left(1+\frac{1}{2}+\dots+\frac{1}{r}-\gamma\right)-\frac{d}{dr}\log n'_r\right\}$$
$$=\tilde{c}_{m,0}\log N-\frac{1}{N^{(m)}}\sum_{r}r^{(m)}n'_r\left(g_r+\frac{d}{dr}\log n'_r\right), \tag{37}$$

where

$$g_r=1+\frac{1}{2}+\dots+\frac{1}{r}-\gamma. \tag{38}$$

In particular, the entropy $c_{1,1}\bumpeq\tilde{c}_{1,1}$, where

$$\tilde{c}_{1,1}=\log N-\frac{1}{N}\sum_{r}rn'_r\left(g_r+\frac{d}{dr}\log n'_r\right). \tag{39}$$

The differentiation can be performed graphically for all r or by numerical differentiation for $r=3, 4, 5, \dots$. (For numerical differentiation see, for example, Jeffreys & Jeffreys (1946, §9·07).) Another estimate of the entropy is $\hat{\tilde{c}}_{1,1}$, where

$$\hat{\tilde{c}}_{1,1}=\log N-\frac{1}{N}\sum_{r}rn_r\left(g_r+\frac{d}{dr}\log n'_r\right), \tag{40}$$

in which the 'prime' has been omitted from the first occurrence of n'_r in (39). This estimate, $\hat{\tilde{c}}_{1,1}$, has leanings towards being an unbiased estimate of the entropy. It can hardly be as good as (39) when the smoothing is reliable. Perhaps the best method of using the present theory for estimating $c_{m,1}$ is to use the compromise $\tilde{c}_{m,1}(t)$ defined in the obvious way by analogy with (33). For large values of r, the factor $g_r + \frac{d}{dr}\log n'_r$ may be replaced by $\log r$ to a good approximation. Terms of $\tilde{c}_{m,1}(t)$ for which this approximation is made, i.e. terms of the form $rn_r \log r$ may be regarded as crude and unadjusted.

7. *Special hypotheses, H.* In this section we shall consider some special classes of hypotheses, H, which determine the distribution p_μ. So far we have taken this distribution as discrete for the sake of logical simplicity. In the present section we shall find it convenient to assume that there is a density function, $f(p)$, where $f(p)\,dp$ is the number of species whose population frequencies lie between p and $p+dp$. (The formulae may of course be generalized to arbitrary distributions by using the Stieltjes integral.) Clearly

$$\int_0^1 f(p)\,dp = s, \tag{41}$$

$$\int_0^1 pf(p)\,dp = 1. \tag{42}$$

The expected value of p for an animal at random from the population is

$$\mathscr{E}(p \mid H) = \int_0^1 p^2 f(p)\,dp = c_{2,0}. \tag{43}$$

The appropriate modifications of the previous formulae are obvious. For example, instead of (10) and (20) we have

$$\mathscr{E}_N(n_r \mid H) = \binom{N}{r}\int_0^1 p^r(1-p)^{N-r} f(p)\,dp, \tag{44}$$

$$\begin{aligned}\mathscr{E}_N(n_r^2 \mid H) = \mathscr{E}_N(n_r \mid H) + \binom{N}{r,\,r}\int_0^1\int_0^1 p^r q^r (1-p-q)^{N-2r} f(p) f(q)\,dp\,dq \\ - \binom{N}{r,\,r}\int_0^1 p^{2r}(1-2p)^{N-2r} f(p)\,dp.\end{aligned} \tag{45}$$

Notice the elegant checks of (44) and (45) that $\mathscr{E}_0(n_0 \mid H) = s$, $\mathscr{E}_1(n_1 \mid H) = 1$, $V_0(n_0 \mid H) = 0$, $V_1(n_1 \mid H) = 0$. Formula (44) leads to the less precise but often more convenient formula

$$\begin{aligned}\mathscr{E}_N(n_r \mid H) &= \frac{1}{r!}\left[1 + O\left(\frac{r^2}{N}\right)\right]\int_0^1 (pN)^r e^{-pN} f(p)\,dp \\ &\simeq \frac{1}{r!}\int_0^1 (pN)^r e^{-pN} f(p)\,dp \quad (r^2 = o(N)),\end{aligned} \tag{46}$$

while a similar treatment of formula (45) leads back merely to formula (22).

We shall now list a number of different types of possible hypotheses and then discuss them. The normalizing constants are all deduced from (42).

H_1 (Pearson's Type I):

$$f(p) = \frac{(\alpha+\beta+2)!}{(\alpha+1)!\,\beta!}\, p^\alpha (1-p)^\beta \quad (\alpha > -1, \beta > -1). \tag{47}$$

H_2 (Pearson's Type III):

$$f(p) = \frac{\beta^{\alpha+2}}{(\alpha+1)!}\, p^\alpha e^{-\beta p} \quad (\alpha > -1, \beta > 0). \tag{48}$$

H_3 (same as H_2 but with $\alpha = -1$):

$$f(p) = \beta p^{-1} e^{-\beta p} \quad (\beta > 0). \tag{49}$$

H_4:

$$f(p) = sAp^{\alpha} \exp(-\beta p - \epsilon p^{-1}) \quad (\beta > 0, \epsilon > 0). \tag{50}$$

H_5 (truncated form of H_3):

$$f(p) = \begin{cases} \beta p^{-1} e^{-\beta p} & (p > p_0), \\ 0 & (p < p_0). \end{cases} \tag{51}$$

H_6 (truncated form of another special case of H_2):

$$f(p) = \begin{cases} \dfrac{p^{-2} e^{-\beta p}}{E(p_0 \beta)} & (p > p_0), \\ 0 & (p < p_0), \end{cases} \tag{52}$$

where $E(w) = -\mathrm{Ei}(-w) = \int_w^{\infty} u^{-1} e^{-u} du$. Ei (w) is known as the 'exponential integral' and has been tabulated several times. (For a list of these tables see Fletcher, Miller & Rosenhead (1946, §§ 13·2 and 13·21).)

We list also a few less completely formulated hypotheses, H_7, H_8 and H_9, for which the population is not explicitly specified, but only the values of $\mathscr{E}_N(n_r \mid H)$. Hence for these hypotheses the parameters may depend on N.

H_7 (Zipf laws):

$$\mathscr{E}(n_r \mid H_7) \propto r^{-\zeta} \quad (r \geqslant 1, \zeta > 0), \tag{53}$$

where ζ is often taken as 2 by Zipf. (See also (94) below.)

H_8 (H_7 with a convergence factor):

$$\mathscr{E}(n_r \mid H_8) = \frac{\lambda x^r}{r^{\zeta}} \quad (r \geqslant 1, \zeta > 0, 0 < x < 1). \tag{54}$$

H_9 (a modification of a special case of H_8):

$$\mathscr{E}(n_r \mid H_9) = \frac{\lambda x^r}{r(r+1)} \quad (r \geqslant 1). \tag{55}$$

We now discuss the nine hypotheses.

(i) H_1 has the advantage that the exact formula (44) can be evaluated in elementary terms. We can see from (41) and (43) that

$$s = \frac{\alpha + \beta + 2}{\alpha + 1}, \tag{56}$$

$$\mathscr{E}(p \mid H_1) = \frac{\alpha + 2}{\alpha + \beta + 3} = \frac{(\alpha + 2)(\alpha + \beta + 2)}{(\alpha + 1)(\alpha + \beta + 3)} \frac{1}{s}. \tag{57}$$

In most applications we want $f(p)$ to be small when p is not small and $\mathscr{E}(p \mid H)$ to be large compared with $1/s$. Hence if a hypothesis of the form H_1 is to be appropriate at all, we shall usually want β to be large, by (47), and α to be close to -1, by (57).

By (44) we see that

$$\mathscr{E}_N(n_r \mid H_1) = \binom{N}{r} \frac{(\alpha + \beta + 2)!\,(\alpha + r)!\,(\beta + N - r)!}{(\alpha + 1)!\,\beta!\,(\alpha + \beta + N + 1)!}. \tag{58}$$

Hence, by (2′), if the smoothed values n'_r and n'_{r+1} were equal to their expectations, given H_1, we would have

$$r^* = \frac{(\alpha + r + 1)(N - r)}{\beta + N - r}. \tag{59}$$

(ii) H_2 can be regarded as a convenient approximation to H_1 if $\beta > 0$. Strictly, the hypothesis H_2 is impossible since it allows values of p greater than 1, but it gives all such values of p combined a very small probability provided that β is large. H_2 was used by Greenwood & Yule (1920) and by Fisher (see Corbet *et al.* 1943). We have

$$s = \frac{\beta}{\alpha+1}, \quad \mathscr{E}(p \mid H_2) = \frac{\alpha+2}{\beta} = \frac{\alpha+2}{\alpha+1}\frac{1}{s}, \tag{60}$$

so that α must be close to -1. Hence, if $r^2 = o(N)$,

$$\mathscr{E}_N(n_r \mid H_2) \simeq \frac{\beta(\alpha+r)!}{(\alpha+1)!\,r!}\left(\frac{N}{N+\beta}\right)^r \left(\frac{\beta}{N}\right)^{\alpha+1}, \tag{61}$$

which is of the negative binomial form.

(iii) Of all hypotheses of the form H_2, Fisher (Corbet *et al.* 1943) was mainly concerned with H_3, the case $\alpha = -1$. (See example (i) in § 8 below.) Then

$$s = \infty, \quad \mathscr{E}(p \mid H_3) = \frac{1}{\beta}, \tag{62}$$

$$\mathscr{E}_N(n_r \mid H_3) \simeq \frac{\beta}{r}\left(\frac{N}{N+\beta}\right)^r = \frac{\beta x^r}{r} \quad (r^2 = o(N)), \tag{63}$$

say. For large samples, x (which, unlike β, depends on N) is close to 1 and the factor x^r may be regarded as a convergence factor which prevents $\sum_{r=1}^{\infty} \mathscr{E}_N(n_r \mid H_3)$ from becoming infinite. The convergence factor also increases the likelihood of being able to find a satisfactory fit to given frequencies, n_r, merely because it involves a new parameter.

We see from (22) that

$$V_N(n_r \mid H_3) \simeq \mathscr{E}_N(n_r \mid H_3)\left\{1 - \frac{1}{2^{2r+1}}\binom{2r}{r}\left(\frac{N}{N+\beta}\right)^r\right\}. \tag{64}$$

If $\beta r = o(N)$ it follows that

$$\frac{V_N(n_r \mid H_3)}{\mathscr{E}_N(n_r \mid H_3)} \simeq 1 - \frac{1}{2^{2r+1}}\binom{2r}{r}. \tag{65}$$

Thus in these circumstances $V_N(n_r \mid H_3)$ lies between the bounds given by (26) and (27), being for each r about twice as close to the smaller bound than to the larger one. When applying the chi-squared test, where χ^2 is defined by equation (19), we can hardly go far wrong by assuming (65) to be applicable whatever the distribution determined by H may be. But, of course, we may often be able to improve on (65) when H is specified in terms of the distribution of p. For convenience in applying (65) we give a short table of values of

$$k_r = \left[1 - \frac{1}{2^{r+1}}\binom{2r}{r}\right]^{-1} \tag{66}$$

r	k_r	r	k_r	r	k_r
1	1·33	6	1·13	11	1·10
2	1·23	7	1·12	12	1·09
3	1·19	8	1·11	13	1·09
4	1·16	9	1·11	14	1·08
5	1·14	10	1·10	15	1·08

For larger values of r, the approximation $1 + 1/\{(2\sqrt{(\pi r)})\}$ is correct to two places of decimals.

Suppose we are given a sample of size N and we wish to estimate β and x. The method used by Fisher was to equate the observed values of $\Sigma r n_r = N$ and $\Sigma n_r = S$ to their expected values. (Note that S is the observed number of species and should not be confused with s.) This led him to the equations

$$-\beta \log_e(1-x) = S, \quad N = \beta x/(1-x), \tag{67}$$

$$\frac{N}{S} = \frac{x}{-(1-x)\log_e(1-x)}, \tag{68}$$

which he solved by using a table of $x/(1-x)$ in terms of $\log_{10}(N/S)$.

A theoretically more satisfactory method of estimating β and x would be by minimizing χ^2, defined by (19), with $r = \infty$. This method leads to equations which would be most laborious to solve by hand but which will be given here since large-scale computers now exist. To prevent misunderstanding we mention at once that Fisher obtained a perfectly good fit by the simpler method, in his example, i.e. example (i) of §8 below, though, as pointed out in §8, H_3 must not be too literally regarded as true.

By (65) we may write

$$\chi^2 = \sum_{r-1}^{\infty} k_r \left(\frac{\beta x^r}{r} - 2n_r + \frac{r n_r^2}{\beta x^r}\right). \tag{69}$$

The equations giving β and x will then be

$$\beta^2 \Sigma k_r x^r / r = \Sigma r k_r n_r^2 x^{-r}, \tag{70}$$

$$\beta^2 \Sigma k_r x^r = \Sigma r^2 k_r n_r x^{-r}, \tag{71}$$

and these equations could be solved iteratively.

When β and x are specified the cumulative sums of $\mathscr{E}_N(n_r \mid H_3)$ can be found by making use of the approximation

$$\sum_t^{t \geqslant r} \frac{x^t}{t} \simeq E(-r \log_e x) + \frac{x^r}{2r}\left(1 + \tfrac{1}{6}\log_e x - \frac{1}{6r}\right), \tag{72}$$

which will be a very good approximation if the terms involving $\frac{1}{6}\log x$ and $\frac{1}{6r}$ are negligible. This approximation can be obtained by means of the Euler-Maclaurin summation formula. (See, for example, Whittaker & Watson (1935, §7·21).)

(iv) We have just seen that when $\alpha = -1$ in H_2 we obtain $s = \infty$ and of course $\mathscr{E}_N(n_0 \mid H) = \infty$. There are strong indications in examples (ii), (iii) and (iv) of §8 that we may wish to take $\alpha < -1$, and then even worse divergencies occur. For example, if $\alpha = -2$ we would obtain, from (61), the intolerable result

$$\mathscr{E}_N(n_1 \mid H_2)/\mathscr{E}_N(n_2 \mid H_2) = \infty.$$

In order to avoid these divergencies we could in theory use hypothesis H_4, with a small value of ϵ. Unfortunately, this hypothesis seems to be analytically unwieldy; it is mentioned partly for its interest as intermediate between Pearson's Types III and V.

(v) Another method of avoiding divergencies is to use truncated distributions. These truncated distributions are not theoretically pleasing but at least some of them can be handled analytically. H_5 is a truncated form of H_3. We may describe p_0 as the smallest possible population frequency of any species. In most applications it would be difficult to obtain a sample large enough to determine p_0 with any accuracy. In fact if the estimate of

p_0 were to be reliable the sample would need to be so large that n_r would vanish for all small values of r. In the examples of §8, n_1 is always larger than any other value of n_r, so these samples would need to be increased greatly before one could expect even n_1 to vanish.

We obtain from (41)

$$s = \beta E(p_0\beta). \tag{73}$$

Now

$$E(w) = -\gamma - \log_e w + w - \frac{w^2}{2!\,2} + \frac{w^3}{3!\,3} - \dots, \tag{74}$$

an equation which is undoubtedly well known. It can be proved, for example, by using Dirichlet's formula for γ. (See, for example, Whittaker & Watson (1935, § 12·3, example 2).) In particular, if w is very small,

$$E(w) \sim -\log_e(\gamma' w), \tag{75}$$

where

$$\gamma' = e^{\gamma} = 1{\cdot}781072. \tag{76}$$

(Cf. Jahnke & Emde (1933, p. 79), where our γ' is denoted by γ.) Since p_0 is assumed to be small, we have

$$s \bumpeq -\beta \log(p_0\gamma'\beta), \quad p_0 \bumpeq \beta^{-1} e^{-\gamma - s/\beta}. \tag{77}$$

On applying equation (46) we see that

$$\mathscr{E}_N(n_r \mid H_5) \bumpeq \frac{\beta}{r}\left(\frac{N}{N+\beta}\right)^r \quad (r \geqslant 1, r^2 = o(N)), \tag{78}$$

$$\mathscr{E}_N(n_0 \mid H_5) \bumpeq \beta E[p_0(N+\beta)] \bumpeq -\beta \log[\beta_0\gamma'(N+\beta)]. \tag{79}$$

The check may be noticed that equations (77), (78) and (79) are consistent with

$$s = \mathscr{E}(s \mid H_5) = \sum_{r=0}^{\infty} \mathscr{E}_N(n_r \mid H_5).$$

Formula (77) is of some interest, but in most applications both p_0 and s will be largely metaphysical, i.e. observable only within very wide proportional limits.

(vi) The difficulty of determining p_0 would not apply to the same extent if $\alpha = -2$, i.e. for hypothesis H_6. (This hypothesis is fairly appropriate for example (iv) of §8.) We have, by (46),

$$\mathscr{E}_N(n_1 \mid H_6) \bumpeq \lambda x E[p_0(\beta+N)] \bumpeq -\lambda x \log \frac{p_0\gamma' N}{x}, \tag{80}$$

$$\mathscr{E}_N(n_r \mid H_6) \bumpeq \frac{\lambda x^r}{r(r-1)} \quad (r \geqslant 2, r^2 = o(N)), \tag{81}$$

where x and λ, unlike β and p_0, depend on N and are given by

$$x = \frac{N}{N+\beta} \tag{82}$$

and

$$\lambda = \frac{N+\beta}{E(p_0\beta)} \bumpeq -\frac{N+\beta}{\log(p_0\gamma'\beta)}, \quad \beta p_0 \bumpeq \exp\left[-\gamma - \frac{N+\beta}{\lambda}\right]. \tag{83}$$

If λ and x can be estimated from a sample, then β and p_0 can be determined by (82) and (83) and s can then be determined from (41), which gives

$$s + \beta = e^{-p_0\beta}/[p_0 E(p_0\beta)] \bumpeq e^{-p_0\beta}/[-p_0 \log_e(p_0\gamma'\beta)]. \tag{84}$$

In order to estimate λ and x from a sample, one could minimize χ^2, more or less as described above for H_3. For this purpose and for others it may be noted that, by (22),

$$V_N(n_r \mid H_6)/\mathscr{E}_N(n_r \mid H_6) \simeq 1 - \frac{r-1}{2r-1}\frac{1}{2^r}\binom{2r}{r} \quad (r \geqslant 2), \tag{85}$$

$$V_N(n_1 \mid H_6)/\mathscr{E}_N(n_1 \mid H_6) \simeq \frac{\lambda}{N+\beta} \tag{86}$$

$$\simeq \frac{2n_2'}{N} \quad \text{if} \quad x \simeq 1. \tag{87}$$

By comparing (85) with (65) we can get an idea of the smallness of the error arising when calculating χ^2 if (65) is used for hypotheses other than H_3.

Another method of estimating λ and x, rather less efficient, but easier, is the one analogous to that used by Fisher for H_3, namely, we may assume that the expected values of $N-n_1$ and of $S-n_1$ are equal to their observed values, i.e.

$$S-n_1 = \lambda \sum_{r=2}^{\infty} \frac{x^r}{r(r-1)} = \lambda[x+(1-x)\log_e(1-x)] = \lambda(1-e^{-Y}-Y\,e^{-Y}), \tag{88}$$

$$N-n_1 = \lambda \sum_{r=2}^{\infty} \frac{x^r}{r-1} = -\lambda x \log_e(1-x) = \lambda Y(1-e^{-Y}) = \lambda x Y, \tag{89}$$

$$(S-n_1)/(N-n_1) = Y^{-1}-(e^Y-1)^{-1}, \tag{90}$$

where $x = 1-e^{-Y}$. We may solve (90) iteratively, for Y, i.e. $Y = \lim_{n\to\infty} Y_n$, where $Y_1 = 0$ and, for $n = 1, 2, 3, \ldots,$

$$Y_{n+1}^{-1} = \frac{S-n_1}{N-n_1} + (e^{Y_n}-1)^{-1}. \tag{91}$$

When λ and x are specified, the cumulative sums of $\mathscr{E}_N(n_r \mid H_6)$ can be found by making use of the approximation

$$\sum_{t}^{t\geqslant r} \frac{x^t}{t(t-1)} \simeq xE[-(r-1)\log_e x] - E(-r\log_e x) + \frac{x^r}{2r(r-1)}\left(1+\tfrac{1}{6}\log_e x - \frac{1}{3r}\right), \tag{92}$$

which will be a very good approximation if the terms involving $\frac{1}{6}\log x$ and $\frac{1}{3r}$ are negligible (cf. equation (72)). If $(1-x)r$ is small while r is large, then we can prove the following approximation:

$$\sum_{t}^{t\geqslant r} \mathscr{E}_N(n_t \mid H_6) \simeq \lambda x\left\{\frac{1}{r} - (1-x)\,[1-\gamma-\log_e(1-x)-\log_e r]\right\}. \tag{93}$$

If $1-x$ is small but $(1-x)r$ is large, then

$$\sum_{t}^{t\geqslant r} \mathscr{E}_N(n_t \mid H_6) \simeq \frac{\lambda x^r}{(1-x)\,r^2}. \tag{93 A}$$

When in doubt about the accuracy of (93) and (93 A) it is best to use (92), the calculation of which is, however, ill-conditioned, so that the error integrals may be needed to several decimal places.

(vii) We now come to the 'less completely formulated' hypotheses. H_7 is discussed by Zipf, especially with $\zeta = 2$ and also in the slightly modified form

$$\mathscr{E}(n_r \mid H) \propto (r^2 - \tfrac{1}{4})^{-1}. \tag{94}$$

(See Zipf (1949, pp. 546–7), where there are further references, including ones to J. B. Estoup, M. Joos, G. Dewey and E. V. Condon.) Yule (1944, p. 55) refers to Zipf (1932) and objects to Zipf's word distributions on two grounds. First Yule asserts that the fits are unsatisfactory, and secondly he points out that (in our notation)

$$N = \mathscr{E}_N(\Sigma r n_r \mid H_7) = \infty \quad \text{if} \quad 1 < \zeta \leqslant 2,$$

while
$$c_{2,0} = \int_0^1 p^2 f(p)\,dp = \mathscr{E}(p \mid H_7) = \mathscr{E}\left(\frac{\Sigma r(r-1)\,n_r}{N(N-1)} \,\middle|\, H_7\right) = \infty \quad \text{if} \quad 2 < \zeta \leqslant 3.$$

(viii) Yule's second objection to H_7 can be overcome by introducing a 'convergence factor', x^r, giving H_8. If H_7 is any good at all for any particular application then x will be fairly close to 1. It would be of interest to specify H_8 in terms of a density function, $f(p)$, by solving the simultaneous integral equations

$$\frac{\lambda x^r}{r^\zeta} = \frac{1}{r!}\int_0^\infty (Np)^r e^{-Np} f(p)\,dp \quad (r = 1, 2, 3, \ldots). \tag{95}$$

If $\zeta = 1$, then H_8 reduces of course to H_3.

(ix) H_9 is of interest mainly because it works so well in examples (ii) and (iii) of §8. Besides its formal similarity to H_8 with $\zeta = 2$, H_9 also resembles H_6, in virtue of equation (81). A disadvantage of not specifying $f(p)$ is that $V_N(n_r \mid H_9)$ cannot be conveniently worked out from (22), though it can always be estimated from (23) with considerably more work. Moreover, a correct specification of $f(p)$ is more fundamental than that of the expected values of the n_r's and is more likely to lead to a better understanding of the structure of the population.

In order to estimate λ and x from a sample, we could use either of the two methods discussed for H_3 and H_6, except that in the method of minimizing χ^2 it would perhaps be best to guess a formula for $V_N(n_r \mid H_9)$, after experimenting with formula (23). We shall not discuss this method further in this section. The second method consists in determining λ and x from the equations

$$N = \lambda \sum_{r=1}^{\infty} \frac{x^r}{r+1} = -\frac{\lambda}{x}[x + \log_e(1-x)], \tag{96}$$

$$S = \lambda \Sigma \frac{x^r}{r(r+1)} = \frac{\lambda}{x}[x + (1-x)\log_e(1-x)], \tag{97}$$

$$\frac{S}{N} = \frac{x + (1-x)\log_e(1-x)}{-x - \log_e(1-x)}. \tag{98}$$

x can be determined either by tabulating the right-hand side of (98) or by writing $x = 1 - e^{-Y}$ and determining Y from the equation

$$Y^{-1} = (1 - e^{-Y})^{-1} - (1 + S/N)^{-1}. \tag{99}$$

Y can be found iteratively by writing $Y = \lim_{n\to\infty} Y_n$, where $Y_1 = 1+N/S$, and, for $n = 1, 2, 3, \ldots$,

$$Y_{n+1}^{-1} = (1-e^{-Y_n})^{-1} - (1+S/N)^{-1}. \tag{100}$$

Then, by (96), we can find λ from

$$\lambda = \frac{xN}{Y-x}. \tag{101}$$

Having determined λ and x we may wish to test how well H_9 agrees with the sample. For this purpose we need to calculate cumulative sums of the expectations of the n_r's. This can be done by means of the approximation

$$\sum_t^{t\geqslant r} \frac{x^t}{t(t+1)} \simeq E(-r\log_e x) - \frac{1}{x}E[-(r+1)\log_e x] + \frac{x^r}{2r(r+1)}\left(1+\tfrac{1}{6}\log_e x - \frac{1}{3r}\right), \tag{102}$$

deducible from (92). If $(1-x)r$ is small while r is large, then we have the following approximation:

$$\sum_t^{t\geqslant r} \mathscr{E}(n_t \mid H_9) \simeq \lambda x\left\{\frac{1}{r} - (1-x)\left[1-\gamma-\log_e(1-x)-\log_e r\right]\right\}, \tag{103}$$

deducible from and of precisely the same form as (93). An idea of the closeness of this approximation can be obtained from example (ii) below. If $1-x$ is small but $(1-x)r$ is large, then

$$\sum_t^{t\geqslant r} \mathscr{E}(n_t \mid H_9) \simeq \frac{\lambda x^r}{(1-x)r^2}. \tag{103 A}$$

When in doubt about the accuracy of (103) and (103 A) it is best to use equation (102). (See the remarks following equation (93 A).)

8. *Examples*. In each of the four examples given below we use at least two different methods of smoothing the data. One of these methods is, in each example, the graphical smoothing of $\sqrt{n_r}$ for the smaller values of r and another method is the fitting of one or other of the nine special hypotheses of §7. The discussion of these examples is by no means intended to be complete.

Example (i). *Captures of Macrolepidoptera in a light-trap at Rothamsted*. (Summarized from Williams's data (Corbet *et al*. 1943).) $N = 15{,}609$, $S = 240$.

r	n_r	n_r^{iv}	r	n_r	n_r^{iv}	r	n_r	n_r^{iv} (summed)†
1	35	40	11	2	3·5	21–30	18	15·5
2	11	20·0	12	2	3·2	31–50	16	18·0
3	15	13·2	13	5	3·0	51–70	17	11·4
4	14	9·9	14	2	2·8	71–100	8	11·2
5	10	7·9	15	4	2·6	101–150	9	11·8
6	11	6·6	16	3	2·4	151–200	7	7·4
7	5	5·6	17	3	2·3	201–500	12	16·1
8	6	4·8	18	3	2·1	501–1000	6	4·6
9	4	4·3	19	3	2·0	1001–∞	1	0·9
10	4	3·9	20	4	1·9	2349	1	—

† In future tables this word 'summed' will be taken for granted and omitted.

We now present the results of the calculations, followed by comments. (The columns headed n_r^{iv} in the table above are explained in these comments.)

r	n_r	n'_r	n''_r	n'''	n^{iv}_r	r^*	r^{**}	r^{***}	r^{****}
1	35	35	35	35	40	1·1	1·4	1·3	1
2	11	19·4	24·0	22·5	20·0	2·1	2·3	2·2	2
3	15	13·7	18·1	16·3	13·3	3·0	2·9	3·0	3
4	14	10·2	13·1	12·3	10·0	3·8	3·8	3·9	4
5	10	7·8	10·2	9·7	7·9	4·8	4·8	4·8	5
6	11	6·3	8·1	7·7	6·6	5·9	5·9	5·5	6
7	5	5·3	6·8	6·0	5·6	—	—	—	—

The function n'_r was obtained by plotting $\sqrt{n_r}$ against r for $1 \leqslant r \leqslant 20$ and smoothing for $1 \leqslant r \leqslant 7$ by eye, holding in mind the method of least squares. (See note (i) of § 3.) n''_r was obtained in the same way, but an attempt was made to keep away from the graph of n'_r (except at $r = 1$) in order to find out how different a smoothing was reasonable. Next n'''_r was obtained by smoothing the cumulative sums $\sum_{t=1}^{r} tn_t$. Finally, n^{iv}_r is the function obtained by Fisher, i.e. using our hypothesis H_3 (equation (63)) with $\beta = 40{\cdot}2$ and $x = 0{\cdot}9974$. A more complete tabulation of n^{iv}_r is given in the first table. The 'summed' values of n^{iv}_r were calculated by means of equation (72). No statistical test is necessary to see that the fit of n^{iv}_r is very good. The values of r^* corresponding to the four smoothings of the data are denoted by r^*, r^{**}, r^{***} and r^{****} respectively. (Logically this gives r^* two different meanings.) ($r^{****} = 0{\cdot}9974r$, by (2′) and (63).) In accordance with § 3 we could force the r^*'s, etc., to be smooth. This has not been tried here. What is clear is that if H_3 is not accepted then most of the values of r^*, etc., are unreliable to within about 0·2 or 0·3. The approximate values of χ^2 given by (19) with $r = 7$ and assuming (65) are 10·9, 11·1, 9·4 and 11·7 respectively. The number of degrees of freedom is somewhere between 6 and 7. It seems safe to take it as 5 for n'''_r, 6 for n'_r and n''_r and 7 for n^{iv}_r. None of the values of χ^2 is particularly significant, though all are a bit large. The data can be blamed for the largeness of the values of χ^2, since n_2 is obviously much smaller than it ought to be. Of the four smoothings Fisher's seems to be the most likely to give the best approximations to the 'true expectations'. There is hardly anything to choose on the evidence of the sample, but Fisher's smoothing has the advantage of being analytically simple.

The most definite result of interest in this example does not depend much on the smoothing, namely, that the proportion of the population not represented by the species in the sample is about $(35 \pm 5)/15{,}609$. For the '± 5' see formula (65). Perhaps this standard error should be increased slightly, say from 5 to 8, to allow for the preference given to n^{iv}_r.

Formula (77), if it is applicable (i.e. if the truncated form, H_5, of H_3 is assumed), may be written $-\log_{10} p_0 = 1{\cdot}18 + 0{\cdot}011s$, so that if s were say 1000, then the smallest population frequency would be about 10^{-12}. This is mentioned only for its theoretical interest: it is an unjustifiable extrapolation to suppose that the distribution defined by H_5 would stand up to sample sizes large enough to demonstrate clearly the values of s and p_0. N would need to be of the order of $10/p_0$. The proposition which is made probable by the actual sample is that H_3 and H_5 (with the assigned values of the parameters) would give good fits to the values of n_r on other independent samples of 16,000 or less, i.e. that H_3 and H_5 provide good methods of smoothing the data. The cautious tone of this statement can be more fully justified by the following considerations.

If H_5 were reliable then it should be possible to use it to estimate the simpler measures of heterogeneity, such as $c_{2,0}$. Now we can see by (30) that $\hat{c}_{2,0} = 0{\cdot}03935$ and $\hat{c}_{3,0} \simeq 0{\cdot}0035$. (For the calculations, the complete data given by Williams must be used.) Hence, by (30 A), it is reasonable to write $c_{2,0} = 0{\cdot}03935 \pm 0{\cdot}0007$. Let us then see what value for $c_{2,0}$ is implied by H_5. We have

$$\Sigma r(r-1)\, n_r^{\text{iv}} = \beta\Sigma(r-1)\, x^r = \beta x^2/(1-x)^2 = 0{\cdot}0243.$$

$$[\text{As a check,} \quad \int_0^\infty p^2 f(p)\, dp = \beta\int_{p_0}^\infty p\, e^{-\beta p}\, dp = \beta\int_{p_0\beta}^\infty q\, e^{-q} dq \simeq \beta^{-1} = 0{\cdot}025.]$$

Clearly then H_5 cannot be used to estimate $c_{2,0}$. It would be true if misleading to say that H_5 is decisively disproved by the data. Similar remarks would apply in the examples below.

Example (ii). *Eldridge's statistics for fully inflected words in American newspaper English.* Eldridge's statistics (1911) are summarized by Zipf (1949, pp. 64 and 25). We give a summary of Zipf's summary in column (ii) below; more fully in the second table. $N = 43{,}989$, $S = 6{,}001$.

In this example the values of n_r for $r \leqslant 10$ are much larger than in example (i), so we have far more confidence in the smoothing that is independent of particular hypotheses. We shall present some of the numerical calculations in columns and then make comments on each column. We may assert at once, however, by equations (7), (8) and (9), that the proportion of the population represented by the sample is close to $1 - n_1/N = 14/15$. If a foreigner were to learn all 6001 words which occurred in the sample he would afterwards meet a new word at about 6·7 % of words read. If he learnt only $S - n_1 = 3025$ words he would meet a new word about 11·6 % of the time. The corresponding results for word-roots rather than for fully inflected words would be of more interest to a linguist.

(i) r	(ii) n_r	(iii) $\sqrt{n_r}$	(iv) b_r say	(v) $-\Delta$	(vi) rb_r	(vii) rb'_r say	(viii) b'_r	(ix) b'^2_r	(x) n'_r	(xi) r^*
1	2976	54·5	54·5	21·8	54·5	54·5	54·5	2976	2961	0·73
2	1079	32·7	32·7	10·0	65·4	65·4	32·7	1079	1075	1·4
3	516	22·7	22·7	5·7	68·1	67·8	22·6	511	509	2·4
4	294	17·1	17·0	2·8	68·0	70·2	17·5	206	305	3·4
5	212	14·6	14·2	1·8	71·0	72·6	14·5	210	209	4·4
6	151	12·3	12·4	1·5	74·4	74·7	12·4	154	153	5·4
7	105	10·2	10·9	1·3	76·3	76·3	10·9	119	118	6·2
8	84	9·2	9·6	1·2	76·8	76·8	9·6	92	91	—
9	86	9·3	8·4	1·1	75·6	75·6	8·4	71	70	—
10	45	6·7	7·3	—	—	—	—	—	—	—

(i) and (ii). We first consider the values of r only as far as $r = 10$. For larger values of r the smoothing could be done by using k-point smoothing formulae with $k \simeq 2\sqrt{r}$.

(iii) Each entry in this column has standard error of about $\frac{1}{2}$, so one place of decimals is appropriate.

(iv) This column was obtained by smoothing a graph of column (iii) by eye. Experiments with the five-point smoothing formula did not give quite as convincing results. For the five-point smoothing formula, see, for example, Whittaker & Robinson (1944, § 146). For the present application it would be $\sqrt{n'_r} = \sqrt{n_r} - \frac{3}{35}\Delta^4(\sqrt{n_r})$ $(r = 3, 4, 5, \ldots)$.

(v) This column of differences is given as a verification of the smoothness of column (iv). In fact minor adjustments were made in column (iv) in order to improve the smoothness of column (v).

(vi) The numbers b_r of column (iv) are roughly proportional to r^{-1}. This fact suggests that rb_r should be formed and smoothed again in order to improve the smoothing of $\sqrt{n_r}$ still further. This process is of course distinct from assuming that rb'_r should be constant, where the function b'_r is a smoothing of the function b_r.

(vii) and (viii) These columns have already been partly explained. The purpose of this improvement in the smoothing is more for the sake of the ratios n'_{r+1}/n'_r than of the n'_r themselves.

(ix) Where the smoothing of $\sqrt{n_r}$ had no noticeable effect we have taken $b'^2_r = n_r$. It is clearly typical that $b'^2_1 = n_1$, since the eye-smoothing is unlikely to affect n_1 convincingly. Therefore if the smoothing is tested by means of a chi-squared test it will be reasonable to subtract about two degrees of freedom.

(x) We have scaled up column (ix) so as to force $\sum_{r=1}^{9} rn'_r = \sum_{r=1}^{9} rn_r$. We can then assume $N' = N$, convenient for applications of § 6. Note that $\sum_{r=1}^{9} k_r(n'_r - n_r)^2/n'_r = 6{\cdot}5$, so that χ^2, given by (19) and accepting (65) as a good enough approximation, is not significant on eight degrees of freedom. Thus our smoothing is satisfactory, though there may be other satisfactory smoothings.

(xi) r^* is obtained from formula (2′). The larger is r the larger is the standard error of r^*. We may get some idea of the error by means of an alternative smoothing. The standard error of 1^* can be very roughly calculated by an *ad hoc* argument, inapplicable to say 5^*. We may reasonably say that the variance of $2n'_2/n'_1$ with respect to all eye-smoothings will be about the same as that obtained by regarding n'_2 and n'_1 as independent random variables with variances circumscribed by the inequalities (26) and (27), or nearly enough, defined by (65). Now if w and z are independent random variables with expectations W and Z, we have

$$\delta\left(\frac{w}{z}\right) = \frac{\delta w}{Z} - \frac{W\,\delta z}{Z^2},$$

and hence, to a crude approximation,

$$V\left(\frac{w}{z}\right) \simeq \frac{V(w)}{Z^2} + \frac{W^2 V(z)}{Z^4},$$

i.e.

$$\frac{V(w/z)}{W^2/Z^2} \simeq \frac{V(w)}{W^2} + \frac{V(z)}{Z^2}. \tag{104}$$

It follows that

$$\frac{V(1^*)}{1^{*2}} = \frac{V(2n'_2/n'_1)}{(2n'_2/n'_1)^2} \simeq \frac{1}{k_1 n'_1} + \frac{1}{k_2 n'_2}, \tag{105}$$

so that

$$V(1^*) = 0{\cdot}73^2 \times 0{\cdot}0010 = 0{\cdot}00052 \quad \text{and} \quad 1^* = 0{\cdot}73 \pm 0{\cdot}023.$$

(xii) (see the second table). An analytic smoothing which is remarkably good for $r \leq 15$ is given by $n''_r = S/(r^2 + r)$. For larger values of r there is a serious discrepancy, since $\sum_{r=16}^{\infty} n''_r = 374$ while $\sum_{r=16}^{\infty} n_r = 297$. It is clear without reference to the sample that n''_r cannot be satisfactory for sufficiently large values of r, since $\Sigma r n''_r = \infty$ instead of being equal to N.

(i) r	(ii) n_r	(x) n'_r	(xii) n''_r	(xiii) n'''_r	(xi) r^*	(xiv) r^{**}
1	2976	2961	3000	3008	0·73	0·67
2	1079	1075	1000	1002	1·4	1·5
3	516	509	500	500	2·4	2·4
4	294	305	300	301	3·4	3·3
5	212	209	200	201	4·4	4·3
6	151	153	143	144	5·4	5·2
7	105	118	107	108	6·2	6·2
8	84	91	83	84	—	7·2
9	86	70	67	67	—	8·2
10	45	—	55	55	—	—
11–15	156	—	170	170	—	—
16–20	76	—	89	89	—	—
21–30	78	—	92	92	—	—
31–40	34	—	47	47	—	—
41–50	28	—	29	28	—	—
51–60	10	—	19	19	—	—
61–∞	71	—	98	90	—	—
4290	1	—	—	—	—	—

(xiii) The fit can be improved by writing $n'''_r = \lambda x^r/(r^2+r)$ as in equation (55), i.e. using hypothesis H_9. We find by equations (100) and (101) that $\lambda = 6017{\cdot}4$ and $x = 0{\cdot}999667$. Column (xiii) can then be easily calculated directly for $r \leqslant 10$ and by use of (102) or (103) for $r > 10$. ((103) gives the correct values for n'''_9 and n'''_{10}, to the nearest integer, and it gives $\sum\limits_{61}^{\infty} n'''_r = 89{\cdot}96$, as compared with 89·90 when (102) is used.) Note that $\sum\limits_{r=16}^{\infty} n'''_r = 365$, which implies an improvement on n''_r but is still significantly too large. A better fit could be obtained by the method of minimum χ^2 or by using some simple convergence factor other than x^r, such as e^{-ar-br^2} with $a > 0$, $b > 0$.

(xiv) r^{**} is defined as $(r+1)\,n_{r+1}/n_r$ and is equal to $r(r+1)/(r+2)$. This column may be compared with column (xi). The agreement looks fairly good. It is by no means clear which of the two columns gives more reliable estimates of the 'true' values of r^* for $r \leqslant 7$. Column (x) is a better fit to Eldridge's data for $r \leqslant 9$ (and could be extended to be a better fit for all r) than is column (xiii) but is not as smooth. Columns (xii) and (xiii) would be preferable if some theoretical explanation of the analytic forms could be provided. Such an explanation might also show why the fit is not good for large r, even with the convergence factor x. The limitation on r in equation (46) may be relevant.

If H_9 is true, the population parameter $c_{2,0}$, given by (31), can be expressed in the form

$$\frac{\lambda}{N^{(2)}} \sum_{r=1}^{\infty} \frac{r-1}{r+1} x^r = \frac{\lambda}{N^{(2)}} \left[\frac{x}{1-x} - 2 + \frac{2}{x} \log_e (1-x) \right]. \tag{106}$$

Formula (106) would give $c_{2,0} = 0{\cdot}00928$, but this value is probably a bad over-estimate since n'''_r is too large for large r and the terms of $\dfrac{\lambda}{N^{(2)}} \Sigma \dfrac{r-1}{r+1} x^r$ for large r make most of the contribution. Similarly, $\hat{c}_{2,0}$, given by (30), depends mainly on the larger values of r represented in the sample, but Zipf's summary of Eldridge's data is not complete enough to calculate $\hat{c}_{2,0}$. Similarly, assuming H_9, the entropy, $c_{1,1}$, could be estimated from equation

(39), and this method could be expected to give close agreement with the correct value, since $c_{1,1}$ does not depend so very much on the more frequent species. But I have not obtained a closed formula, resembling (106) for example, and the arithmetic required if no closed formula is available would be heavy. The estimation of measures of heterogeneity will be discussed again under example (iii).

Example (iii). *Sample of nouns in Macaulay's essay on Bacon.* (Taken from Yule (1944) Table 4·4, p. 63.) $N = 8045$, $S = 2048$.

r	n_r	r	n_r	r	n_r	r	n_r	r	n_r
1	990	11	24	21	1	31	2	41	1
2	367	12	19	22	4	32	1	45	2
3	173	13	10	23	7	33	1	48	1
4	112	14	10	24	2	34	1	57	1
5	72	15	13	25	1	35	1	58	1
6	47	16	3	26	5	36	1	65	1
7	41	17	10	27	3	37	1	76	1
8	31	18	7	28	4	38	2	81	1
9	34	19	6	29	1	39	4	89	1
10	17	20	5	30	3	40	1	255	1

As in example (ii) we can state some conclusions at once, without doing the smoothing. If our foreigner learns all 2048 nouns that occur in the sample his vocabulary will represent all but (12·3 ± 0·5) % of the population, assuming formulae (9) and (65) or (87). If he learns only 1058 nouns his vocabulary will still represent all but $(n_1 + 2n_2)/N = 19{\cdot}3$ % of the population.

We now present three different smoothings corresponding precisely to those of example (ii).

r	n_r	n'_r	n''_r	n'''_r	r^*	r^{**}	r^{***}	$\frac{d}{dr}\log_{10} n'_r$	$\frac{d}{dr}\log_{10} n'''_r$	$g_r \log_{10} e$
1	990	990	1024	1060	0·74	0·67	0·66	−0·50	−0·65	0·184
2	367	367	341	350	1·4	1·5	1·5	−0·30	−0·37	0·401
3	173	173	170	174	2·6	2·4	2·4	−0·24	−0·26	0·545
4	112	112	102	103	3·4	3·3	3·3	−0·17	−0·20	0·654
5	72	76	68	68	4·4	4·3	4·3	−0·15	−0·16	0·741
6	47	56	49	48	5·3	5·2	5·1	−0·12	−0·14	0·813
7	41	42	35·5	36	6·5	6·2	6·1	−0·11	−0·12	0·876
8	31	34	28·5	28	7·3	7·2	7·1	−0·10	−0·11	0·930
9	34	27	22·7	22	8·2	8·2	8·1	−0·09	−0·10	0·978
10	17	22	18·4	18	—	9·2	9·1	−0·08	−0·09	1·021
11	24	18·5	15·5	15	—	10·2	10·1	—	—	—
12	19	16·0	13·1	12	—	11·1	11·0	—	—	—
13	10	13·7	11·3	10	—	12·1	12·0	—	—	—
14	10	10·9	9·7	9	—	13·1	13·0	—	—	—
15	13	9·6	8·5	8	—	14·1	14·0	—	—	—
16–20	31	32·5	30·5	27	—	—	—	—	—	—
21–30	31	—	31·5	26	—	—	—	—	—	—
31–50	19	—	25·9	19	—	—	—	—	—	—
51–100	6	—	19·9	11	—	—	—	—	—	—
101–∞	1	—	20·3	3·6	—	—	—	—	—	—
255	1	—	—	—	—	254	252	—	—	—

$\sqrt{n'_r}$ was obtained by smoothing $\sqrt{n_r}$ graphically.

$n_r'' = S/(r^2+r)$. It is curious that this should again give such a good fit for values of r that are not too large ($r \leqslant 30$). The sample is of nouns only and, moreover, Yule took different inflexions of the same word as the same.

$n_r''' = \lambda x^r/(r^2+r)$, where $\lambda = 2138{\cdot}90$, $x = 0{\cdot}991074$, the values being obtained from (100) and (101) as in example (ii).

The expressions $\sum_{r=1}^{15} (n_r' - n_r)^2/n_r'$, etc., take the values 9·5, 21·2 and 27·3. The values of χ^2 would be about 2 or 3 larger. (See (19), (26), (27), (65).) There is no question of accepting n_r'' for $r > 50$ but it is better than n_r''' for $r \leqslant 15$. When $r \leqslant 9$ the values of r^* and r^{**} (and therefore of r^{***}) show good agreement except for $r = 1$ and $r = 7$. If the analytic smoothings had not been found, the value of 6^* would have been smoothed off, with repercussions on the function n_r'. The discrepancy in 1^* must be attributed either to a fault in the value of n_1''' (and therefore in H_9) or must be blamed on n_1 (i.e. on sample variation). If I had not noticed the analytic smoothings I would have asserted that $1^* = 0{\cdot}74$ with a standard error of something like 0·04. (See equation (105).)

We now consider two of the measures of heterogeneity in the population, namely, $c_{2,0}$ and $c_{1,1}$. By (30) we can see that $\hat{c}_{2,0} = 0{\cdot}00272$, agreeing with Yule (1944, p. 57). Also $\hat{c}_{3,0} = 0{\cdot}00003957$, so that by (30A) we may reasonably write $c_{2,0} = 0{\cdot}00272 \pm 0{\cdot}00013$. Assuming H_9 to be valid for $r \leqslant 30$, we may also estimate $c_{2,0}$ by $\tilde{c}_{2,0}$ (30) as in equation (33). We have, in a self-explanatory notation,

$$\tilde{c}_{2,0}(30 \mid H_9) = \frac{1}{N^{(2)}}\left\{\lambda \sum_{1}^{30} \frac{r-1}{r+1} x^r + \sum_{30}^{\infty} r^{(2)} n_r\right\}. \tag{107}$$

Now, as in (72),

$$\sum_{31}^{\infty} \frac{r-1}{r+1} x^r \simeq \frac{x^{31}}{1-x} - \frac{2}{x}\left\{E(-32\log_e x) + \frac{x^{32}}{64}\left(1 + \tfrac{1}{6}\log_e x - \tfrac{1}{192}\right)\right\} = 82{\cdot}924.$$

But, as in (106), $\sum_{1}^{\infty} \frac{r-1}{r+1} x^r = 99{\cdot}501$, so that $\sum_{1}^{30} \frac{r-1}{r+1} x^r = 16{\cdot}577$. It follows from (107) that $\tilde{c}_{2,0}(30 \mid H_9) = 0{\cdot}00246$. This is about two standard errors below its expected value, based on the simple unbiased statistic $\hat{c}_{2,0}$. The discrepancy may again be attributed to the large value of n_1'''. If, instead of n_r''', the smoothing n_r'' is accepted for $r \leqslant 30$, we would get $\tilde{c}_{2,0}(30) = 0{\cdot}00267$. (It was in order to obtain this comparison that we calculated $\tilde{c}_{2,0}(30 \mid H_9)$ rather than $\tilde{c}_{2,0}(50 \mid H_9)$. The fit of n_r'' deteriorates at about $r = 30$.)

The last three columns of the table are related to the estimation of the entropy, $c_{1,1}$. (See equation (40) and the remarks following it.) $\frac{d}{dr}\log_{10} n_r'$ was obtained graphically for $r = 1$, 2 and 3 by numerical differentiation for $r = 3, 4, \ldots, 10$. (The graphical and numerical values agreed to two decimal places for $r = 3$.) The column $\frac{d}{dr}\log_{10} n_r'''$ was of course calculated as $\log_{10} x - \left(\frac{1}{r} + \frac{1}{r+1}\right)\log_{10} e$. The crude estimate of the 'entropy to base 10' or 'entropy expressed in decimal digits' is $\log_{10} N - \frac{1}{N}\sum_r r n_r \log_{10} r = 2{\cdot}968$ decimal digits. If n_r' is accepted for $r = 1, 2, 3, \ldots, 10$ we find that

$$\tilde{c}_{1,1}(10) = \log_{10} N - \frac{1}{N}\left\{\sum_{r=1}^{10} r n_r'\left(g_r \log_{10} e + \frac{d}{dr}\log_{10} n_r'\right) + \sum_{r=11}^{\infty} r n_r \log_{10} r\right\} = 3{\cdot}051 \text{ decimal digits.}$$

We shall next calculate $\tilde{c}_{1,1}(50 \mid H_9)$, using another self-explanatory notation. Since, by Jeffreys & Jeffreys (1946, §15·05),

$$g_r \sim \log_e r + \frac{1}{2r} - \frac{1}{12r^2} - \dots,$$

it can be seen that

$$\tilde{c}_{1,1}(50 \mid H_9) \simeq \log_{10} N - \frac{1}{N}\left\{\sum_{r=1}^{10} rn_r''' \left(g_r \log_{10} e + \frac{d}{dr}\log_{10} n_r'''\right)\right.$$
$$\left. + \sum_{11}^{50} rn_r''' \log_{10} r + \log_{10} x \sum_{11}^{50} rn_r''' - \frac{3\log_{10} e}{2}\sum_{11}^{50} n_r''' + \sum_{51}^{\infty} rn_r \log_{10} r\right\}$$
$$= 3{\cdot}192 \text{ decimal digits,}$$

as we may see by means of rather heavy calculations, using the last column of the table, together with equations (72), (74) and (92). The crude estimate of $c_{1,1}$ is the smallest of the three. This is not surprising, since the crude estimate is always too small in the special case of sampling from a population of s species all of which are equally probable.

Example (iv). *Chess openings in games published in the British Chess Magazine*, 1951. For the purposes of this example we arbitrarily regard the openings of two games as equivalent only if the first six moves (three white and three black) are the same and in the same order in both games. $N = 385$, $S = 174$.

r	n_r	n_r'	n_r''	n_r'''	r^{**}
1	126	126	126	126	0·39
2	22	22	24·6	24	1·0
3	5	7·6	8·1	8	2·0
4	4	4·8	4·0	4	3·0
5	3	3·2	2·4	2·4	4·0
6	4	2·6	1·6	1·6	5·0
7	0	2·2	1·1	1·14	6·0
8	3	—	0·85	0·86	7·0
9	1	—	0·66	0·67	8·0
10	1	—	0·52	0·53	9·0
11	0		3·97	4·80	
13	1				
14	1				
16	1				
23	1				
36	1				
∞	—				

$\sqrt{n_r'}$ was obtained by graphical smoothing of $\sqrt{n_r}$.

n_r'' was obtained by assuming H_6 (see equation (52)), i.e. $n_r'' = \mathscr{E}_{385}(n_r \mid H_6)$, where the parameters x and λ were obtained from (91) and (89). These gave $x = 0{\cdot}99473$, $\lambda = 49{\cdot}635$ and n_r'' for $r \geqslant 2$ is then given by (81). Next p_0 was determined as $0{\cdot}00011304 = 1/8846$ by using equation (80). Then (82) gave $\beta = 2{\cdot}040$, so that, in accordance with (52) and (74),

$$f(p) \simeq \begin{cases} 0{\cdot}128p^2 e^{-2{\cdot}040p} & (p > 1/8846), \\ 0 & (p < 1/8846). \end{cases}$$

Finally, equation (84) gives $s = 1132$. This then is the estimate of the total number of openings in the population, though the sample is too small to put any reliance in it.

$$n_r'''(r \geqslant 2) \text{ is simply } (S-n_1)/(r^2-r) = 48/(r^2-r).$$

This is just as good a fit as n_r''. It gives an infinite value to $c_{2,0}$, but this is not as serious an objection as it sounds since H_6 would also give quite the wrong value for $c_{2,0}$. (Cf. the concluding remarks in the discussion of example (i).)

We list in the table the values of r^* corresponding to n_r'', calling the values r^{**} in conformity with the convention of the present section. Clearly $r^{**} = (r-1)x$ when $r \geqslant 2$. Thus the average population frequency of the 126 openings that each occurred once only in the sample is $0{\cdot}39/385 = 0{\cdot}001$.

A player who learnt all 174 openings would expect to recognize about 67 % of future openings for the same population, assuming that the sample was random. If he learnt the 48 openings that each occurred twice or more in the sample the percentage would drop to 55 % and if he learnt the 26 that occurred three times or more the percentage would drop to 49 %. (See formula (6′).)

9. *Index of notations having a fixed meaning.*

§1. N, n_r (but see also §2), n_0, q_r, r^* (as a definition of the asterisk, but there is a slight change of convention in §8), n_r' (here again there is a slight change in §8), $\mathscr{E}(\ )$, $V(\ )$.

§2. $s, p_\mu, H(p_1, p_2, \ldots, p_s) = H, \mathscr{E}_N, \mu'_{r,t,N}$.

§3. N'.

§5. $x_{\mu,r} = x_\mu, a_\mu$.

§6. $c_{m,n}, \hat{c}_{m,0}, \tilde{c}_{m,0}, \tilde{c}_{m,0}(t), \gamma, g_r, \hat{\tilde{c}}_{1,1}, \tilde{c}_{m,1}(t)$.

§7. $p, f(p), p_0, H_1$ to $H_9, E(\ ), k_r, S, \gamma'$.

REFERENCES

ANSCOMBE, F. J. (1948). The transformation of Poisson, binomial and negative binomial data. *Biometrika*, **35**, 246–54.

ANSCOMBE, F. J. (1950). Sampling theory of the negative binomial and logarithmic series distributions. *Biometrika*, **37**, 358–82.

BARTLETT, M. S. (1936). The square root transformation in the analysis of variance. *J. R. Statist. Soc. Suppl.* **3**, 68–78.

CHAMBERS, E. G. & YULE, G. U. (1942). Theory and observation in the investigation of accident causation. (Including discussion by J. O. Irwin and M. Greenwood.) *J. R. Statist. Soc. Suppl.* **7**, 89–109.

CORBET, A. S., FISHER, R. A. & WILLIAMS, C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* **12**, 42–58.

ELDRIDGE, R. C. (1911). *Six Thousand Common English Words*. Buffalo: The Clements Press. (Mentioned in Zipf (1949).)

FLETCHER, A., MILLER, J. C. P & ROSENHEAD, L. (1946). *An Index of Mathematical Tables*. London: Scientific Computing Service.

GOOD, I. J. (1950*a*). A proof of Liapounoff's inequality. *Proc. Camb. Phil. Soc.* **46**, 353.

GOOD, I. J. (1950*b*). *Probability and the Weighing of Evidence*. London: Charles Griffin.

GOODMAN, L. A. (1949). On the estimation of the number of classes in a population. *Ann. Math. Statist.* **20**, 572–9.

GREENWOOD, M. & YULE, G. U. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *J. R. Statist. Soc.* **83**, 255–79.

HARDY, G. H. (1949). *Divergent Series*. Oxford: Clarendon Press.

JAHNKE, E. & EMDE, F. (1933). *Funktionentafeln mit Formeln und Kurven*, 2nd ed. Leipzig and Berlin.

JEFFREYS, H. (1948). *Theory of Probability*, 2nd ed. Oxford: Clarendon Press.

JEFFREYS, H. & JEFFREYS, B. S. (1946). *Methods of Mathematical Physics*. Cambridge University Press.

JOHNSON, W. E. (1932). Appendix (edited by R. B. Braithwaite) to 'Probability: deductive and inductive problems'. *Mind*, **41**, 421–3.

NEWBOLD, E. M. (1927). Practical applications of the statistics of repeated events, particularly to industrial accidents. (Including discussion by M. Greenwood, D. R. Wilson, M. Culpin, E. Farmer and L. Isserlis.) *J. R. Statist. Soc.* **90**, 487–547 (esp. Appendix, pp. 518–35).

PRESTON, F. W. (1948). The commonness, and rarity, of species. *Ecology*, **29**, 254–83.

SHANNON, C. E. (1948). A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **27**, 379–423.

SIMPSON, E. H. (1949). Measurement of diversity. *Nature, Lond.*, **163**, 688.

USPENSKY, J. V. (1937). *Introduction to Mathematical Probability*. New York: McGraw Hill.

WHITTAKER, E. T. & ROBINSON, G. (1944). *The Calculus of Observations*, 4th ed. London and Glasgow: Blackie.

WHITTAKER, E. T. & WATSON, G. N. (1935). *A course of Modern Analysis*, 4th ed. Cambridge University Press.

YULE, G. U. (1944). *Statistical Study of Literary Vocabulary*. Cambridge University Press.

ZIPF, G. K. (1932). *Selected Studies of the Principle of Relative Frequency in Language*. Harvard University Press. (Mentioned in Yule (1944) and Zipf (1949.)

ZIPF, G. K. (1949). *Human Behaviour and the Principle of Least Effort*. Cambridge, Mass.: Addison-Wesley Press.

CAPTURE-RECAPTURE ANALYSIS

By J. M. HAMMERSLEY

Lectureship in the Design and Analysis of Scientific Experiment, University of Oxford

This paper presents a new method for the analysis of capture-recapture data. Previous methods have employed deterministic or partially stochastic models of the population under study; the new method not only provides a fully stochastic model, but it also allows death-rates to depend upon both current time and the health and age of the separate individuals. The numerical analysis associated with the method is, however, heavy, and can only be recommended when one wants to extract the maximum amount of information from data which have required much effort or time to collect. The method is illustrated by an estimation of the death-rate of Alpine Swifts (*Apus melba*) in the wild state. Some thirty years having been spent in collecting these data, it does not seem disproportionate to have spent several months in computing the results. The larger part of this calculation was done on ordinary desk calculators.

Some unusual problems in maximum-likelihood estimation are discussed.

INTRODUCTION

The literature upon capture-recapture analysis is now extensive.* To acquaint himself with it, the reader may consult the references given at the end of this paper, and then follow up the references cited in turn in these papers. To make the present paper fairly self-contained, however, we begin with a brief outline of the problem.

The main object is to estimate the (possibly changing) size of a population of individuals (e.g. animals, birds, fish, insects, etc.) in their natural state. To this end, the experimenter captures individuals from the population on a number of successive occasions. On each occasion the catch is to be considered as a random sample of individuals from the population; that is to say, each individual in the population (if alive) has an equal chance of being captured on any given occasion irrespective of its age, health, type, etc., and of any previous captures it may have suffered, although this chance may vary from one occasion to another. Each time an individual is caught, a record is marked for it or on it to show the occasion of that capture; the individual is then returned to the population. Each individual in the population may die at any time, the chance of death being supposed to depend only upon the current time, and the current age, type and condition of the individual. The experimenter has to estimate the size of the population at any given instant from records of the several occasions of capture. The main difficulty is that this estimation depends upon an estimate of the effective death-rate then prevailing in the population, and that this latter estimate is hard to make.

The experimenter may or may not be able to distinguish between one individual and another. For instance, in the case of birds, he can place a numbered ring upon the leg of each captured bird; and, if all rings bear distinct numbers, he will be able to identify any particular bird upon every occasion that it is subsequently recaptured. On the other hand, in the case of moths, the marks may be only spots of paint on the wing, the colour and position of the marks indicating the dates of capture but failing to discriminate between two moths which have both suffered the same history of captures. Merely for convenience in describing the theory, we shall suppose that the experimenter has fastened individually numbered rings upon all individuals at their first capture; but we shall later see that as a rule the analysis can still be carried through in the other case.

---

* Laplace had ideas on the subject 170 years ago, and the last 60 years have produced a steady flow of papers.

Let us then suppose that the experimenter captures individuals at a sequence of discrete instants $T_1, T_2, \ldots, T_m$; and that the individuals which are captured at some time or another are named $I_1, I_2, \ldots, I_n$ by means of their distinct rings numbered $1, 2, \ldots, n$. The complete set of data then takes the form of an $m \times n$ matrix in which the element in row i and column j is 1 or 0 according as the individual I_j was or was not captured (or recaptured) at the instant T_i. Call this data matrix $\mathbf{D} = (d_{ij})_{mn}$.

There are two ways of analysing **D**. According to the first scheme, which we call 'analysis by rows', we first consider any given subset of rows of **D**, and count the number of columns which contain 1 in every row of the subset and zero in every row not belonging to the subset. Let us call this the α-number of the subset. We then analyse the set of all α-numbers obtained by considering every possible subset of rows. This method leads to very considerable algebraic difficulties since the α-numbers so obtained are not independent of each other, and therefore cannot be combined in a simple fashion to yield the likelihood of **D**. Analysis by rows is basically the structure which underlies most previous attacks on the capture-recapture problem; but, either because of the algebraic difficulties just mentioned or because of limitations in the data presented to them, previous authors have adopted one or more of the following simplifying expedients:

(*a*) Certain subsets of rows are grouped together according to some rule of procedure, and the analysis is confined to the totals of the α-numbers falling in the groups thus generated. This pooling of α-numbers may perhaps sacrifice some information. Also it will be dictated to some extent should the available data not give complete records of individual histories.

(*b*) No use is made of the α-numbers corresponding to subsets which comprise a single row only; that is to say, we disregard those individuals which are captured only once and never recaptured. This sacrifices information because these special α-numbers influence the estimates of the experimenter's catching efficiency, on the precision of which estimates depends also the precision of the death-rate estimate. Again this will be dictated if the data does not contain records of individuals that are never recaptured.

(*c*) It is assumed that the total catch at any given instant of catching is a small fraction of the population at risk of capture; and that therefore the distribution of the α-numbers, which is strictly a multihypergeometric distribution, may be replaced by a multinomial distribution.

(*d*) It is assumed that death-rates operate deterministically; that is to say, the proportion of birds, having any given history of captures, which will die in a specified period of time is not subject to random variation.

(*e*) It is assumed that death-rate is independent of age and current time.

The alternative scheme is 'analysis by columns'. This simple change of viewpoint (which, for the analysis of the Alpine Swifts, is in a sense to take a bird's-eye view of the data!) sweeps away all the algebraic difficulties and allows us to avoid, if we so desire and if the comprehensiveness of the data permits, any or all of the palliatives (*a*)–(*e*) above. Although analysis by columns has these algebraic advantages, it has, on the other hand, the disadvantage that it leads (as might be expected) to more complicated equations and considerably heavier computing than in the case when analysis by rows is undertaken with one or more of the assumptions (*a*)–(*e*). Analysis by columns involves finding separately for each column of **D** the likelihood of the observations in that column. Then the likelihood of **D** is the product of these separate likelihoods; for the histories of two individuals are mutually independent under the assumptions made in the second paragraph of this introduction. The most serious restriction in these assumptions is probably that the chance of death depends only upon current time and current age, type and condition of the individual; for in reality the death of one individual may affect the chances of another's survival, as in case of over-population.

One other artifice, which will be fundamental, is the

THEOREM. *Parameter spaces may be transformed before or after maximum-likelihood estimation with the same final result in either case.*

This is certainly not an original statement, and it is quite obvious once made. It is, of course, false for most forms of estimation other than maximum likelihood. Statisticians, to whom I have mentioned it, fall into two classes of roughly equal size: those of one class knew it from their cradles, those of the other show momentary disbelief. I would have belonged to the second class had not this capture-recapture analysis forced the theorem to my notice.

The analysis has thrown up a further aspect of maximum-likelihood estimation. Suppose that the likelihood L depends upon a parameter $\boldsymbol{\theta}$ with co-ordinates $\theta_1, \theta_2, \ldots, \theta_q$. Then as a rule the maximum-likelihood solution is obtained as the solution of the simultaneous equations

$$\partial L/\partial\theta_j = 0 \quad (j = 1, 2, \ldots, q). \tag{1}$$

What should one do, however, if no solution exists for these equations, or (worse) if the equations themselves do not exist (for example, the likelihood function may not be differentiable with respect to θ_j)? Even the rigorous text-books seem to overlook this possibility; and yet it may arise in the very simplest of practical problems. For instance, suppose that we wish to estimate the parameter π of a binomial distribution having performed N trials of which n are 'successful'. The likelihood function is

$$L = \binom{N}{n} \pi^n (1-\pi)^{N-n},$$

and hence

$$\frac{1}{L}\frac{\partial L}{\partial \pi} = \begin{cases} -N/(1-\pi) & (n=0), \\ (n-N\pi)/\pi(1-\pi) & (0<n<N), \\ N/\pi & (n=N). \end{cases}$$

When $0<n<N$, we get the usual solution $\pi = n/N$; but when $n = 0$ or N, the equation $\partial L/\partial\pi = 0$ has no solution. In these cases it seems reasonable to take $\pi = 0$ and $\pi = 1$ respectively; for, subject to the known information $0 \leqslant \pi \leqslant 1$, these values of π make the likelihood as large as possible. It is easy, however, to construct examples in which no value of the parameter makes the likelihood as large as possible. For instance, consider the foregoing binomial distribution, when $n = 0$ and we have the prior restriction $0 < \pi$. In general the following procedure may be acceptable to some readers, and I, for one, would accept it for want of a better. Let $\boldsymbol{\theta}$ be an unknown vector parameter restricted to a known set Θ, and let $\overline{\Theta}$ denote the closure of Θ. Since the likelihood function satisfies $0 \leqslant L \leqslant 1$, it possesses a supremum $\bar{L}$ for fixed data and $\boldsymbol{\theta}$ belonging to Θ. Let $\epsilon > 0$ be prescribed and let $\overline{\Theta}(\epsilon)$ denote the closure of the set of all $\boldsymbol{\theta}$ in Θ such that $L \geqslant \bar{L} - \epsilon$ for the fixed data. Then

$\overline{\Theta}(\epsilon_1) \supseteq \overline{\Theta}(\epsilon_2)$ whenever $\epsilon_1 \geqslant \epsilon_2$; and hence $\overline{\Theta}(0) = \lim_{\epsilon \to 0+} \overline{\Theta}(\epsilon)$ exists in $\overline{\Theta}$. It is to be noticed that $\overline{\Theta}(0)$ may not have any points in Θ, it may not be a unique point or even a countably infinite set of points, and that the likelihood for any $\boldsymbol{\theta}$ in $\overline{\Theta}(0)$ may be less than every L for $\boldsymbol{\theta}$ in Θ. It is, indeed, possible to construct an example in which the likelihood is not constant and yet, for almost all data, every point of the parameter space is simultaneously a maximum-likelihood and a minimum-likelihood solution; for example, it suffices to take L equal to a positive constant or zero according as θ is rational or not, except that L is to be suitably modified for one particular value of the data (thus preventing a solution which is independent of the data). Nevertheless, I should be prepared to define $\overline{\Theta}(0)$ as the maximum-likelihood solution, mainly on the grounds that it reduces to the conventional maximum-likelihood solution whenever this latter exists, and that in many other cases it seems to produce the sensible answer. The problem now remains of specifying some expression for the variance of this estimate; but I have no satisfactory idea how to set about this. In the first place one has to define what may be meant by the variance of an uncountable statistic. Secondly, even when the statistic is unique, one has to determine what should replace the conventional expression $\partial^2 \log L/\partial\theta_i\,\partial\theta_j$. Presumably if one can calculate a variance-covariance matrix $\mathbf{V}$ for any unique $\boldsymbol{\theta}$, then what is wanted in general is an expression of the type $M\mathbf{V}$, where M is a generalized averaging over $\overline{\Theta}(0)$. Thus, if $\overline{\Theta}(0)$ is measurable and not of zero measure, M might denote normalized integration with respect to this measure. Alternatively, one might prefer an averaging defined as the extraction of a supremum or infimum; and a number of other possibilities will suggest themselves to the reader. The only precaution to observe is that the averaging shall be regular in the sense that it is an identity mapping for a point estimate $\hat{\boldsymbol{\theta}}$; for then the definition of the variance-covariance matrix will include the conventional definition. There is a further point to watch. Suppose equations (1) are soluble in the ordinary sense except for $j = 1$. If we now estimate θ_1 by the foregoing general technique and substitute this solution into the remaining equations, then some or all of these remaining equations may no longer be soluble. We may expect this to happen when the estimate lies on the frontier of parameter space; and we shall have an instance of it when we come to consider the data for the Alpine Swifts.

A STOCHASTIC MODEL FOR CAPTURE-RECAPTURE ANALYSIS

Let $p_i = 1 - q_i$ denote the probability that a particular individual I_j, known to be alive on occasion T_i, is captured (or recaptured) on that occasion. Since (by hypothesis) this probability is constant for all individuals at risk of capture, p_i is an index of the experimenter's catching efficiency on the occasion T_i. Let $\phi_{ij} = 1 - \psi_{ij}$ denote the probability that the individual I_j is alive on the occasion T_{i+1}, given that it is alive on occasion T_i. Let μ_{ij} denote the probability that the individual I_j is recaught on some occasion subsequent to T_i, given that it is captured (or recaptured) on occasion T_i. Then we have the fundamental difference equation

$$\mu_{ij} = \phi_{ij}(p_{i+1} + q_{i+1}\mu_{i+1,j}); \quad \mu_{mj} = 0. \tag{2}$$

The boundary condition $\mu_{mj} = 0$ holds because the experimenter makes no further catches after the final occasion T_m. For a proof of the difference equation itself, suppose we know that I_j has been caught on occasion T_i. Then we know it is alive on occasion T_i. If it is recaught, it must certainly survive until occasion T_{i+1} at least, of which event there is probability ϕ_{ij}; and then it must either be recaught on occasion T_{i+1} given that it is then alive (with probability p_{i+1}), or it must escape capture on occasion T_{i+1} (with probability q_{i+1}) and then be captured on some subsequent occasion given that it is alive at T_{i+1} (and the probability of this is $\mu_{i+1,j}$). The 'either-or' of the previous sentence is exclusive; and (2) follows at once. Now from (2)

$$\phi_{ij} = \phi_{ij}(p_{i+1}+q_{i+1}) \geqslant \phi_{ij}(p_{i+1}+q_{i+1}\mu_{i+1,j}) = \mu_{ij}$$

and

$$\mu_{ij} = \phi_{ij}\mu_{i+1,j}+\phi_{ij}p_{i+1}(1-\mu_{i+1,j}) \geqslant \phi_{ij}\mu_{i+1,j},$$

because the various probabilities lie between 0 and 1 inclusive. By putting $p_{i+1} = 1$ and 0 respectively in these inequalities we deduce that

$$\phi_{ij} \geqslant \mu_{ij} \geqslant \phi_{ij}\mu_{i+1,j} \tag{3}$$

is the best possible set of inequalities operating upon the μ's. Hence (3) defines the parameter space $\boldsymbol{\mu}$. The parameter space $\boldsymbol{\phi}$ is evidently the closed unit hypercube. From (2) we also have

$$\mu_{i-1,j}-\phi_{i-1,j} = \phi_{i-1,j}(p_i-1+q_i\mu_{ij}) = \phi_{i-1,j}q_i(-1+\mu_{ij});$$

and so

$$\phi_{i-1,j}q_i = \frac{\phi_{i-1,j}-\mu_{i-1,j}}{1-\mu_{ij}} \quad (i = 2, 3, \ldots, m). \tag{4}$$

Also

$$\frac{p_i}{q_i} = \frac{\phi_{i-1,j}-\phi_{i-1,j}q_i}{\phi_{i-1,j}q_i} = \frac{\phi_{i-1,j}(1-\mu_{ij})-(\phi_{i-1,j}-\mu_{i-1,j})}{\phi_{i-1,j}-\mu_{i-1,j}}$$

$$= \frac{\mu_{i-1,j}-\phi_{i-1,j}\mu_{ij}}{\phi_{i-1,j}-\mu_{i-1\,j}} \quad (i = 2, 3, \ldots, m). \tag{5}$$

Suppose I_j is captured on occasions $T_\alpha, T_\beta, \ldots, T_\theta$ and on no other occasions, and that we consider only information upon I_j on occasions from T_α onwards. The likelihood is then

$$L_j = \phi_{\alpha j}\phi_{\alpha+1,j}\cdots\phi_{\theta-1,j}p_\alpha p_\beta\cdots p_\theta\frac{q_\alpha q_{\alpha+1}\cdots q_\theta}{q_\alpha q_\beta\cdots q_\theta}(1-\mu_{\theta j}). \tag{6}$$

The likelihood of **D** will now be

$$L = L_1 L_2 \ldots L_n. \tag{7}$$

If we wish to take into account the possibility that I_j has remained uncaught in the population from some unknown instant of entry $K(j)$, then L_j must be multiplied by

$$\phi_{K(j),j}\phi_{K(j)+1,j}\cdots\phi_{\alpha-1,j}q_{K(j)}q_{K(j)+1}\cdots q_{\alpha-1},$$

in which $K(j)$ is a parameter to be estimated (or a random variable to be integrated out).

The maximum-likelihood solution results from choosing, subject to (3), the available parameters so that L is as large as possible. Before we can do this, however, we must by some assumption reduce the number of unknown parameters; for at present there are more unknown parameters than observations (i.e. elements of **D**), and hence the parameters cannot be uniquely determined. The analysis will, moreover, not be tractable from the numerical point of view unless we keep the number of unknown parameters fairly small (say less than 100). Thus we might assume that all individuals were equally healthy. At the same time, also for numerical convenience, we may employ the theorem stated in the introduction. According to this we may transform the parameter space before maximizing (7), and then, after estimates have been found in this transformed space, we can transform them back into the original space. We shall adopt this technique in the illustrations on Alpine Swifts by employing (4) and (5) to carry **p**-space into $\boldsymbol{\mu}$-space. Another technique worth considering is the deliberate use of an excess number of parameters, by the introduction of new parameters which depend in a functionally determinate manner upon the minimal set of parameters assumed in formulating the model. In the augmented parameter space the minimal set will be some hypersurface. We may then maximize (7) upon this hypersurface by the use of Lagrangian undetermined multipliers. This will prove a convenient procedure when it is easier to work with many simple equations than with a smaller number of complicated ones.

The parameters which are sought are the ϕ's and the p's. The former specify death-rate, and the latter lead immediately to estimates of population size.

AN EXPERIMENT WITH ALPINE SWIFTS

We shall analyse some data collected upon two colonies of Alpine Swifts (*Apus melba*) at Solothurn in Switzerland from 1920 to 1950.

A bird aged x years is called a nestling if $x < 1$, young if $1 \leqslant x < 2$, and adult if $x \geqslant 2$. An ornithologist can tell the difference between these three categories upon inspection. Alpine Swifts migrate annually. If a bird returns to the place of catching when it is adult, it is ornithologically reasonable to suppose that it will thereafter continue to return to this same place each year until it dies. This does not necessarily hold for nestlings or young. When a bird is caught for the first time a ring is placed upon its leg. All rings bear distinct numbers. Only a small number of birds are caught when young; and it is convenient, in view of the anomalous migration of young, to suppress all records of capturing young. A bird is said to be 'ringed' when it is caught for the first time in the *unsuppressed* records. We shall suppose that an adult bird has a constant probability $\phi = 1-\psi$ of surviving one year. This is a much simpler assumption than that of the previous general treatment. It is not an unfair supposition in view of the fact that most birds die due to natural hazards rather than through old age. These hazards vary from year to year, for meteorological reasons amongst others; but the estimate, which is wanted, is naturally a hypothetical one referring to an 'average' bird in an 'average' year. Both these last two uses of the word 'average' are defined by the assumption that ϕ is to be regarded as a constant; that is to say, the death-rate for Alpine Swifts is the metric upon which an average is taken rather than, say, the mean winter temperatures or the supply of summer food. Since ϕ no longer depends upon a suffix j we may write μ_i for the probability that an adult bird (known to be alive in year i) is subsequently recaught. To deal with the anomalous behaviour of nestlings and young, we write π for the probability that a nestling (ringed at birth) survives *and* returns to the place of catching *two* years later. Define $\omega = \pi/\phi$.

We can now write down the probabilities for the various possible records on a single bird:

[1] Bird ringed in the unsuppressed records as an adult in year i and not recaught:

$$L_j = (1-\mu_i). \tag{8}$$

[2] Bird ringed in the unsuppressed records as an adult in year i and recaught in years $\alpha, \beta, \ldots, \theta$ and in no other years:

$$L_j = \phi^{\theta-i} p_\alpha p_\beta \cdots p_\theta \frac{q_{i+1} q_{i+2} \cdots q_\theta}{q_\alpha q_\beta \cdots q_\theta}(1-\mu_\theta). \tag{9}$$

[3] Bird ringed as a nestling in year $i-1$ and not recaught in the unsuppressed records:

$$\begin{aligned} L_j = (1-\pi) + \pi q_{i+1}(1-\mu_{i+1}) &= 1-\phi\omega+\phi\omega(1-p_{i+1}-q_{i+1}\mu_{i+1}) \\ &= 1-\omega\mu_i, \end{aligned} \tag{10}$$

by virtue of (2).

[4] Bird ringed as a nestling in year $i-1$ and in the unsuppressed records recaught in years $\alpha, \beta, \ldots, \theta$ and in no other years:

$$L_j = \text{expression (9) multiplied by } \omega. \tag{11}$$

[5] Bird found dead in year θ: in expression (9) we must replace $\phi(1-\mu_\theta)$ by ψ. Hence it is enough to introduce the factor

$$\frac{1-\phi}{\phi(1-\mu_\theta)}. \tag{12}$$

[6] Bird killed artificially for experimental purposes in year θ: introduce the factor

$$\frac{1}{1-\mu_\theta}. \tag{13}$$

To simplify the ensuing numerical analysis we now transform from the parameter space $(\phi, \pi, \mathbf{p})$ to the space $(\phi, \omega, \boldsymbol{\mu})$. This is justified by the theorem quoted in the introduction and effected by substituting (4) and (5) into (9) to yield

$$\begin{aligned}
&\phi^{\theta-i} p_\alpha p_\beta \cdots p_\theta \frac{q_{i+1} q_{i+2} \cdots q_\theta}{q_\alpha q_\beta \cdots q_\theta} (1-\mu_\theta) \\
&= \left(\frac{\phi-\mu_i}{1-\mu_{i+1}}\right)\left(\frac{\phi-\mu_{i+1}}{1-\mu_{i+2}}\right) \cdots \left(\frac{\phi-\mu_{\theta-1}}{1-\mu_\theta}\right)(1-\mu_\theta)\left(\frac{\mu_{\alpha-1}-\phi\mu_\alpha}{\phi-\mu_{\alpha-1}}\right)\left(\frac{\mu_{\beta-1}-\phi\mu_\beta}{\phi-\mu_{\beta-1}}\right) \cdots \left(\frac{\mu_{\theta-1}-\phi\mu_\theta}{\phi-\mu_{\theta-1}}\right) \\
&= (\phi-\mu_i)\left(\frac{\phi-\mu_{i+1}}{1-\mu_{i+1}}\right)\left(\frac{\phi-\mu_{i+2}}{1-\mu_{i+2}}\right) \cdots \left(\frac{\phi-\mu_{\theta-1}}{1-\mu_{\theta-1}}\right)\left(\frac{\mu_{\alpha-1}-\phi\mu_\alpha}{\phi-\mu_{\alpha-1}}\right)\left(\frac{\mu_{\beta-1}-\phi\mu_\beta}{\phi-\mu_{\beta-1}}\right) \cdots \left(\frac{\mu_{\theta-1}-\phi\mu_\theta}{\phi-\mu_{\theta-1}}\right).
\end{aligned} \tag{14}$$

Upon noting that $\mu_m = 0$ and collecting up expressions (8)–(14) inclusive, we find for the likelihood of the whole set of data $\mathbf{D}$,

$$L = \omega^A \left(\frac{1-\phi}{\phi}\right)^F \prod_{i=1}^{m-1} \{(1-\omega\mu_i)^{B_i}(1-\mu_i)^{C_i}(\phi-\mu_i)^{D_i}(\mu_i-\phi\mu_{i+1})^{E_i}\}, \tag{15}$$

where in the unsuppressed records

A = number of nestlings which are recaptured;
B_i = number of nestlings ringed in year $i-1$ which are never recaptured;
C_i = (number of adults ringed in year i and never recaptured)
  − (number of adults ringed before year i and recaptured after year i)
  − (number of nestlings ringed before year $i-1$ and recaptured after year i)
  − (number of birds killed artificially or recaptured dead in year i);
D_i = number of birds which, having been ringed *either* as adults before year $i+1$ *or* as young before year i, were recaptured sometime after year $i+1$ but which were not recaptured in year $i+1$;
E_i = number of birds recaptured in year $i+1$;
F = number of birds found dead but not artificially killed.

It it thus seen that, under suitable assumptions, it is not necessary to know individual histories.

Let us renumber the suffices i so that they run over the values $i = 20, 21, \ldots, 50$ corresponding to the years 1920–50 inclusive. The data then yields $A = 205$, $F = 6$ and

i	B_i	C_i	D_i	E_i
20	0	+ 1	2	0
21	17	0	2	0
22	18	− 2	6	1
23	0	− 5	4	4
24	12	0	7	7
25	33	+ 2	3	7
26	25	− 5	11	0
27	12	0	11	0
28	35	− 11	17	0
29	25	− 17	17	7
30	0	− 21	15	7
31	27	− 17	10	18
32	24	− 23	18	18
33	42	− 29	28	25
34	60	− 37	31	28
35	82	− 51	34	42
36	100	− 51	26	58
37	129	− 15	44	16
38	103	− 41	51	9
39	82	− 51	57	3
40	129	− 58	47	17
41	178	− 57	41	21
42	125	− 53	36	37
43	175	− 44	60	23
44	113	− 65	51	33
45	132	− 66	60	43
46	155	− 90	61	61
47	189	− 96	59	81
48	216	− 103	42	87
49	3	− 58	0	84
50	328	+ 18	0	0

(16)

The summary (16) shows that the terms in L involving μ_{20} are

$$(1-\mu_{20})(\phi-\mu_{20})^2.$$

To maximize L we must therefore minimize μ_{20}. According to (3) we get

$$\mu_{20} = \phi\mu_{21}. \tag{17}$$

Now substitute (17) into (15) and pick out the terms in μ_{21}

$$(1-\omega\mu_{21})^{17}(\phi-\mu_{21})^2(1-\phi\mu_{21})(\phi-\phi\mu_{21})^2.$$

Once again use (3) to obtain from (17)

$$\mu_{20} = \phi\mu_{21} = \phi^2\mu_{22}. \tag{18}$$

We can now eliminate μ_{20} and μ_{21} from (15). The terms in μ_{22} become

$$\phi^4(1-\phi\omega\mu_{22})^{17}(1-\omega\mu_{22})^{18}(\phi-\mu_{22})^6(1-\phi\mu_{22})^2(1-\phi^2\mu_{22})(\mu_{22}-\phi\mu_{23}). \tag{19}$$

The analysis may proceed in a straightforward manner until we reach the next run of zeros in E_i. Treating these in a similar fashion we find

$$\mu_{26} = \phi\mu_{27} = \phi^2\mu_{28} = \phi^3\mu_{29}. \tag{20}$$

Upon elimination of μ_{26}, μ_{27} and μ_{28} from (15), the stage is set for logarithmic differentiation, and the estimation follows from the simultaneous solution, subject to (3), of

$$\frac{E_i}{\mu_i-\phi\mu_{i+1}} = \frac{\omega B_i}{1-\omega\mu_i}+\frac{C_i}{1-\mu_i}+\frac{D_i}{\phi-\mu_i}+\frac{\phi E_{i-1}}{\mu_{i-1}-\phi\mu_i} \quad (i = 23, 24, 30, 31, \ldots, 48, 49), \tag{21}$$

$$\frac{1}{\mu_{22}-\phi\mu_{23}} = \frac{18}{1-\omega\mu_{22}}+\frac{17\phi}{1-\phi\omega\mu_{22}}+\frac{6}{\phi-\mu_{22}}+\frac{2\phi}{1-\phi\mu_{22}}+\frac{\phi^2}{1-\phi^2\mu_{22}}, \tag{22}$$

$$\frac{7}{\mu_{25}-\phi^4\mu_{29}} = \frac{33}{1-\omega\mu_{25}}+\frac{3}{\phi-\mu_{25}}+\frac{2}{1-\mu_{25}}+\frac{7\phi}{\mu_{24}-\phi\mu_{25}}, \tag{23}$$

$$\begin{aligned}\frac{7}{\mu_{29}-\phi\mu_{30}} = {}&\frac{25\omega}{1-\omega\mu_{29}}+\frac{35\phi\omega}{1-\phi\omega\mu_{29}}+\frac{12\phi^2\omega}{1-\phi^2\omega\mu_{29}}+\frac{25\phi^3}{1-\phi^3\omega\mu_{29}}\\ &+\frac{17}{\phi-\mu_{29}}+\frac{11\phi^2}{1-\phi^2\mu_{29}}-\frac{5\phi^3}{1-\phi^3\mu_{29}}+\frac{7\phi^4}{\mu_{25}-\phi^4\mu_{29}},\end{aligned} \tag{24}$$

$$\mu_{50} = 0, \tag{25}$$

$$\begin{aligned}0 = \Omega \equiv \frac{\partial \log L}{\partial \omega} = {}&\frac{205}{\omega}+\sum_i{}'\frac{-B_i\mu_i}{1-\omega\mu_i}\\ &-\frac{17\phi\mu_{22}}{1-\phi\omega\mu_{22}}-\frac{18\mu_{22}}{1-\omega\mu_{22}}-\frac{33\mu_{25}}{1-\omega\mu_{25}}\\ &-\frac{25\phi^3\mu_{29}}{1-\phi^3\omega\mu_{29}}-\frac{12\phi^2\mu_{29}}{1-\phi^2\omega\mu_{29}}-\frac{35\phi\mu_{29}}{1-\phi\omega\mu_{29}}-\frac{25\mu_{29}}{1-\omega\mu_{29}},\end{aligned} \tag{26}$$

$$\begin{aligned}0 = \Phi \equiv \frac{\partial \log L}{\partial \phi} = {}&\sum_i{}'\left\{\frac{D_i}{\phi-\mu_i}-\frac{E_i\mu_{i+1}}{\mu_i-\phi\mu_{i+1}}\right\}+\frac{37}{\phi}-\frac{6}{1-\phi}\\ &-\frac{17\omega\mu_{22}}{1-\phi\omega\mu_{22}}+\frac{6}{\phi-\mu_{22}}-\frac{2\mu_{22}}{1-\phi\mu_{22}}-\frac{2\phi\mu_{22}}{1-\phi^2\mu_{22}}\\ &+\frac{3}{\phi-\mu_{25}}-\frac{28\phi^3\mu_{29}}{\mu_{25}-\phi^4\mu_{29}}-\frac{\mu_{23}}{\mu_{22}-\phi\mu_{23}}-\frac{35\omega\mu_{29}}{1-\phi\omega\mu_{29}}\\ &-\frac{24\phi\omega\mu_{29}}{1-\phi^2\omega\mu_{29}}-\frac{75\phi^2\omega\mu_{29}}{1-\phi^3\omega\mu_{29}}+\frac{17}{\phi-\mu_{29}}\\ &-\frac{22\phi\mu_{29}}{1-\phi^2\mu_{29}}+\frac{15\phi^2\mu_{29}}{1-\phi^3\mu_{29}}-\frac{7\mu_{30}}{\mu_{29}-\phi\mu_{30}}.\end{aligned} \tag{27}$$

In equations (26) and (27), $\sum_i{}'$ denotes summation over $i = 23, 24, 30, 31, \ldots, 48, 49$. The method of solving the foregoing set of 28 simultaneous equations will be discussed in the next section.

The next stage is the determination of the inverse variance-covariance matrix for $(\phi, \omega, \boldsymbol{\mu})$. As already mentioned in the introduction, it is not clear what one should do with co-ordinates on the frontier of the parameter space. In default of a better procedure, I have regarded (18) and (20) as algebraic identities; and thus taken

$$p_{21} = p_{22} = p_{27} = p_{28} = p_{29} = 0$$

as quantities not subject to sampling error. It then remains to calculate

$$c_{ij} = \frac{\partial}{\partial \mu_i}\left(\frac{1}{L}\frac{\partial L}{\partial \mu_j}\right) \quad (i,j = 22, 23, 24, 25, 29, 30, \ldots, 48, 49);$$

$$c_{i\phi} = \frac{\partial}{\partial \mu_i}\left(\frac{1}{L}\frac{\partial L}{\partial \phi}\right), \quad c_{i\omega} = \frac{\partial}{\partial \mu_i}\left(\frac{1}{L}\frac{\partial L}{\partial \omega}\right) \quad (i = 22, 23, 24, 25, 29, 30, \ldots, 48, 49);$$

$$c_{\phi\omega} = \frac{\partial}{\partial \phi}\left(\frac{1}{L}\frac{\partial L}{\partial \omega}\right), \quad c_{\phi\phi} = \frac{\partial}{\partial \phi}\left(\frac{1}{L}\frac{\partial L}{\partial \phi}\right), \quad c_{\omega\omega} = \frac{\partial}{\partial \omega}\left(\frac{1}{L}\frac{\partial L}{\partial \omega}\right).$$

The algebraic expressions for these quantities are lengthy but quite easily obtained from L, and therefore will not be printed here. The only non-vanishing c's are c_{ii}, $c_{\phi\phi}$, $c_{\omega\omega}$, $c_{\omega\phi} = c_{\phi\omega}$, $c_{i,i+1} = c_{i+1,i}$, $c_{i\omega} = c_{\omega i}$, $c_{i\phi} = c_{\phi i}$.

If we write

$$\mathbf{C} = \left(\begin{array}{c:cc} c_{ij} & c_{i\omega} & c_{i\phi} \\ \hdashline c_{\omega j} & c_{\omega\omega} & c_{\omega\phi} \\ c_{\phi j} & c_{\phi\omega} & c_{\phi\phi} \end{array}\right) \quad (i,j = 22, 23, 24, 25, 29, 30, \ldots, 49);$$

then $-\mathbf{C}^{-1}$ is the variance-covariance matrix of the estimates in the parameter space $(\boldsymbol{\mu}, \omega, \phi)$. To transform to the $(\mathbf{p}, \omega, \phi)$ space we take a first variation of (4) and obtain

$$\delta q_i = -\frac{1}{\phi(1-\mu_i)}\,\delta\mu_{i-1} + \frac{(\phi-\mu_{i-1})}{\phi(1-\mu_i)^2}\,\delta\mu_i + \frac{\mu_{i-1}}{\phi^2(1-\mu_i)}\,\delta\phi,$$

with the exception

$$\delta q_{26} = -\frac{1}{\phi(1-\phi^3\mu_{29})}\,\delta\mu_{25} + \frac{\phi^2(\phi-\mu_{25})}{(1-\phi^3\mu_{29})}\,\delta\mu_{29} + \left\{\frac{\mu_{25}}{\phi^2(1-\phi^3\mu_{29})} + \frac{3\phi\mu_{29}(\phi-\mu_{25})}{(1-\phi^3\mu_{29})^2}\right\}\delta\phi.$$

Consider now the 26×27 matrix $\mathbf{U}$ whose rows and columns are indexed according to the following scheme:

$$\begin{array}{c|cccc:ccccc:cc} & 22 & 23 & 24 & 25 & 29 & 30 & \cdots & 48 & 49 & \omega & \phi \\ 23 & \times & \times & & & & & & & & & \times \\ 24 & & \times & \times & & & & & & & & \times \\ 25 & & & \times & \times & & & & & & & \times \\ 26 & & & & \times & \times & & & & & & \times \\ \hdashline 30 & & & & & \times & \times & & & & & \times \\ 31 & & & & & & \times & \ddots & & & & \times \\ \vdots & & & & & & & \ddots & \ddots & & & \vdots \\ 48 & & & & & & & & \times & & & \times \\ 49 & & & & & & & & \times & \times & & \times \\ \hdashline \omega & & & & & & & & & \times & \times & \times \\ \phi & & & & & & & & & & \times & \times \end{array} \tag{28}$$

and whose elements vanish except in the last column and on the leading diagonal and super-diagonal. These remaining elements are specified by

$$u_{i+1,i} = -1/\phi(1-\mu_i) \quad (i = 22, 23, 24, 29, 30, \ldots, 47, 48);$$

$$\left.\begin{aligned} u_{ii} &= (\phi-\mu_{i-1})/\phi(1-\mu_i)^2 \\ u_{i\phi} &= \mu_{i-1}/\phi^2(1-\mu_i) \end{aligned}\right\} \quad (i = 23, 24, 25, 30, 31, \ldots, 48, 49);$$

$$u_{26,25} = -1/\phi(1-\phi^3\mu_{29});$$

$$u_{26,29} = \phi^2(\phi-\mu_{25})/(1-\phi^3\mu_{29})^2;$$

$$u_{26,\phi} = \frac{\mu_{25}}{\phi^2(1-\phi^3\mu_{29})} + \frac{3\phi\mu_{29}(\phi-\mu_{25})}{(1-\phi^3\mu_{29})^2};$$

$$u_{\omega\omega} = u_{\phi\phi} = 1; \quad u_{\omega,49} = u_{\omega\phi} = u_{\phi\omega} = 0.$$

Then the variance-covariance matrix for estimates in the $(\mathbf{p}, \omega, \phi)$ space is $-\mathbf{UC}^{-1}\mathbf{U}'$, whose rows and columns are indexed in the same fashion as the rows of $\mathbf{U}$ in (28).

The estimates of $\mathbf{p}$ themselves follow from (4).

NUMERICAL ANALYSIS

The main numerical labour lies in the solution of the 28 simultaneous equations (21)–(27) inclusive. This was done by iteration as follows:

(i) Choose trial values of ϕ, ω and μ_{22}. Then use equations (21)–(24) inclusive to calculate in succession $\mu_{23}, \mu_{24}, \mu_{25}, \mu_{29}, \mu_{30}, \ldots, \mu_{49}, \mu_{50}$, this last value of μ_{50} coming from (21) with $i = 49$. This sequence of calculations will be called a minor cycle. The final value of μ_{50} so obtained should be zero to agree with (25); but in fact it will not be, due to the inexact trial choice of μ_{22}. So keeping ϕ, ω both fixed, adjust the trial value of μ_{22} and begin a new minor cycle. Repeat the minor cycles with successive adjustments of μ_{22} until the resulting value of μ_{50} satisfies (25).

Then

(ii) use equations (26) and (27) to calculate Ω and Φ. The work so far covered in (i) and (ii) will be called a major cycle.

(iii) Do two more major cycles with fresh trial values of ϕ, ω. We now have three pairs of values of Ω, Φ. Carry out a bivariate inverse interpolation in the (Ω, Φ)-plane to estimate values of ϕ, ω such that $\Omega = \Phi = 0$.

(iv) Finally, repeat the major cycles, with successive approximations upon ϕ, ω until (26) and (27) are satisfied.

In carrying out the minor cycles we have to observe (3). Thus any minor cycle can be stopped, when only partly completed, as soon as (3) is violated by a poor choice of μ_{22}. This helps to shorten the work considerably. Increasing μ_{22} increases all the other μ's for fixed ϕ, ω; and decreasing μ_{22} decreases all the other μ's. The latter values of μ are extremely sensitive to small changes in μ_{22}; in the final solution a change of 10^{-10} in μ_{22} gave a change of 4×10^{-5} in μ_{50}. Similarly, in the major cycles Ω and Φ were very sensitive to changes in μ_{50}. On that account the work was carried out retaining 10 significant figures throughout, with a number of interpolating refinements built into the final stages. After the final major cycle, the computing errors in ϕ and ω were determined variationally and found to be 4×10^{-6}

and 1×10^{-6} respectively. This leaves sufficient margin in view of the standard errors of these quantities.

All the foregoing work was done on ordinary desk calculators. A minor cycle could be completed in about a day; and a major cycle in about 10 days.

The 27×27 matrix $\mathbf{C}$ was evaluated on a desk calculator. $\mathbf{C}$ was inverted on SEAC. This required a couple of days for coding, followed by about 4 hr. machine time. In view of the large proportion of zeros in C, however, this inversion could probably have been accomplished with about two months' work on a desk calculator. Finally, the standard errors were calculated from the diagonal elements of $-\mathbf{U}\mathbf{C}^{-1}\mathbf{U}'$ on a desk calculator.

NUMERICAL RESULTS

The final major cycle employed the values

$$\phi = 0{\cdot}82177\,252, \quad \omega = 0{\cdot}15231\,687.$$

i	μ
22	0·49114 70078
23	0·55283 78505
24	0·48551 39880
25	0·39790 15014
29	0·57753 15026
30	0·62797 49077
31	0·68978 94273
32	0·65635 69119
33	0·64690 45524
34	0·63783 00155
35	0·63643 78806
36	0·57590 81349
37	0·37055 55369
38	0·37884 51461
39	0·42295 19018
40	0·50268 84610
41	0·54322 89445
42	0·57591 99689
43	0·54504 16713
44	0·59087 04335
45	0·61829 34809
46	0·64147 85385
47	0·63248 97872
48	0·57021 21461
49	0·40731 92325
50	0·00000 00000

these values being interpolated between the independent minor cycles

$$\mu_{22} = 0{\cdot}49114\,70079 \text{ leading to } \mu_{50} = +0{\cdot}00002\,70661,$$

$$\mu_{22} = 0{\cdot}49114\,70078 \text{ leading to } \mu_{50} = -0{\cdot}00001\,02407.$$

For a reason to be noticed presently, we classify all birds as either X-birds or Y-birds. A bird is an X-bird if and only if it is a nestling when ringed, and is a Y-bird if and only if it is an adult when ringed. Since ringing, by definition, refers only to the unsuppressed records, these two classes are mutually exclusive and exhaustive. The following table summarizes for each year the number of birds captured, together with estimates of the experimenter's catching efficiency and of the current population size:

Year	Number of birds caught			Catching efficiency (p_i ± standard error)	Population size (± standard error)
	X-birds	Y-birds	All birds		
1920	0	3	3	—	—
1921	17	2	19	0	—
1922	23	0	23	0	—
1923	2	2	4	0·100 ± 0·096	40 ± 38
1924	19	11	30	0·364 ± 0·154	82 ± 35
1925	36	16	52	0·320 ± 0·105	162 ± 53
1926	32	6	38	0·241 ± 0·084	158 ± 55
1927	12	11	23	0	—
1928	37	4	41	0	—
1929	33	0	33	0	—
1930	7	3	10	0·201 ± 0·071	50 ± 18
1931	38	7	45	0·240 ± 0·082	188 ± 64
1932	38	17	55	0·533 ± 0·094	103 ± 18
1933	64	20	84	0·430 ± 0·080	195 ± 36
1934	82	25	107	0·412 ± 0·065	260 ± 41
1935	113	22	135	0·384 ± 0·060	351 ± 54
1936	141	35	176	0·468 ± 0·056	376 ± 45
1937	170	64	234	0·525 ± 0·052	446 ± 45
1938	121	17	138	0·116 ± 0·028	1190 ± 285
1939	91	9	100	0·066 ± 0·022	1515 ± 494
1940	136	2	138	0·024 ± 0·014	5726 ± 3279
1941	202	8	210	0·150 ± 0·035	1401 ± 326
1942	143	25	168	0·201 ± 0·041	837 ± 173
1943	213	42	255	0·342 ± 0·050	745 ± 109
1944	139	16	155	0·177 ± 0·035	876 ± 172
1945	172	30	202	0·264 ± 0·042	765 ± 121
1946	191	39	230	0·309 ± 0·041	743 ± 100
1947	236	60	296	0·403 ± 0·043	734 ± 78
1948	271	58	329	0·464 ± 0·041	709 ± 63
1949	47	72	119	0·484 ± 0·043	246 ± 22
1950	285	55	340	—	—

The expectation of life of a bird is $1/(1-\phi)$ years, and this, together with its standard error, is 5·61 ± 0·30 years.

The probability that a bird will survive 2 years from birth and then return to its place of birth is $\phi\omega$, and this, together with its standard error, is 0·1252 ± 0·0088. Thus about one nestling in eight returns two years after birth.

COMMENT ON THE DATA AND THEIR NUMERICAL RESULTS

The data were collected by H. Arn of Solothurn, who was primarily interested in quite different problems to those for which his data have now been used. In particular, because of his other interests, he made a deliberate attempt in some years to catch unringed adults and in other years to ring as many nestlings as possible. As a result the catching is not random, the various birds at risk have not always had equal chances of capture, and the fundamental assumption of the present capture-recapture method is not satisfied. The table above shows evidence of this lack of random catching. For example, in 1949 the experimenter appears to have concentrated on capturing adult birds; and in this year he has secured an unusually large number of Y-birds in a fairly small total catch. As a result in 1949 his catching efficiency (based upon the false assumption that he had been catching at random) is artificially inflated, and therefore the estimate of population size is made much too low. The reverse effect takes place in the years 1939, 1940 and 1941 when he has evidently concentrated on ringing nestlings.

It was nevertheless decided to analyse the data as though they had been based upon random catching, because considerable ornithological interest attaches to the age of Alpine Swifts in that they enjoy an unusually long expectation of life in the wild state, and because the present data is quite unique in its extensiveness. Since the assumption of random catching is not here valid, the numerical results must be interpreted with a good deal of reserve. The individual year-by-year population sizes carry very little weight, although the general tendency of the size of the population to increase with time is borne out by other evidence not presented here. The probability of 1/8 for the return of a nestling two years later is also distinctly suspect. On the other hand, the estimate of expectation of life is in good accord with estimates obtained from other less extensive experiments, and is less liable to be incorrect although its standard error may well be wrong.

The fact that the numerical results are for these reasons suspect will concern the ornithologist: but the main object of the present paper is to provide a theoretical analysis of the general capture-recapture problem, and to show that the analysis can in fact be handled numerically. This has been demonstrated.

I am indebted to Herr H. Arn of Solothurn, who collected the original data upon Alpine Swifts, and to Dr D. Lack of the Edward Grey Institute of Field Ornithology, to whom the data were communicated and by whom they were transmitted to me. I am also indebted to the computing staff of the Lectureship in the Design and Analysis of Scientific Experiment, University of Oxford, for carrying out much of the numerical analysis. In particular, I am glad to thank Mrs M. Parke who carried the main burden of the work. I also wish to thank Dr J. Curtiss, Mr J. Todd, and the staff of the Computation Laboratory of the National Bureau of Standards, Washington, D.C., for the use of SEAC to invert the 27 × 27 reciprocal variance-covariance matrix. I should like to record helpful discussions on the subject of capture-recapture analysis with Dr D. J. Finney, Prof. R. A. Fisher, Mr P. H. Leslie, and Prof. P. A. P. Moran. Finally, I am indebted to the referee for useful comments.

REFERENCES

BAILEY, N. T. J. (1951). On estimating the size of mobile populations from recapture data. *Biometrika*, **38**, 293–306.

CHAPMAN, D. G. (1952). Inverse, multiple and sequential sample censuses. *Biometrics*, **8**, 286–306.

LESLIE, P. H. & CHITTY, D. (1951). The estimation of population parameters from data obtained by means of the capture-recapture method. I. The maximum likelihood equations for estimating the death-rate. *Biometrika*, **38**, 269–92.

LESLIE, P. H. & CHITTY, D. (1952). The estimation of population parameters from data obtained by means of the capture-recapture method. II. The estimation of total numbers. *Biometrika*, **39**, 363–88.

MORAN, P. A. P. (1951). A mathematical theory of animal trapping. *Biometrika*, **38**, 307–11.

MORAN, P. A. P. (1952). The estimation of death-rates from capture-mark-recapture sampling. *Biometrika*, **39**, 181–8.

THE USE OF CHAIN-BINOMIALS WITH A VARIABLE CHANCE OF INFECTION FOR THE ANALYSIS OF INTRA-HOUSEHOLD EPIDEMICS

BY NORMAN T. J. BAILEY

Nuffield Lodge, Regent's Park, London

1. INTRODUCTION

An important characteristic of highly infectious diseases is that they tend to produce groups of cases occurring within households rather than separate cases scattered throughout the community. If all the cases in a household were due to some single source of infection such as typhoid-infected water, then we should expect the total number of cases to follow a binomial distribution. On the other hand, if the disease were introduced into the household by one of its members and then transmitted from person to person, a very different situation would arise. For a disease involving a short period of high infectivity and an approximately constant incubation period, we should expect to be able to distinguish different generations of the intra-household epidemic. There might be a binomial distribution of secondary cases resulting from contact with the primary case, to be followed later by another binomial distribution of tertiary cases amongst the susceptibles who had previously escaped, and so on. In his classic paper on this subject Greenwood (1931) introduced such a chain-binomial model for the investigation of measles epidemics, for which the assumptions made are thought to be approximately true. When the period of infectivity is more extended, as with scarlet fever or whooping cough for example, the stochastic model recently discussed by Bailey (1953) may be more appropriate. Greenwood examined data on the 1926 measles epidemic in St Pancras, and showed that, so far as the total number of cases in a household was concerned, the hypothesis of a simple binomial distribution was quite inadequate, while the distribution expected from the chain-binomial model gave a satisfactory fit to the observations. It is important to notice, however, that this material provided information only on the *total* size of the epidemic in a given household; no analysis of the individual links of the chain was possible. Now Wilson, Bennett, Allen & Worcester (1939), in their investigation of cases of measles occurring in Providence, Rhode Island, during 1929–34, were able to go a stage further and break down the data into its constituent parts. Of course it must be recognized that departures from a constant incubation period, or the possibility of multiple primary cases, give rise to certain difficulties in the so-called chaining of this kind of material. However, it seemed likely that in the present case this source of confusion would not be serious. Wilson *et al.* were able to show that in none of the groups of available data did the Greenwood model give an adequate description of the numbers of cases occurring in the separate generations through which the epidemic passed, though it did sometimes give a satisfactory fit to the distribution of the total number of cases in a household. In a later paper, Wilson (1947) considered a somewhat different approach. According to this, in families containing two susceptibles in addition to the first case, for example, one could estimate separately the chance of one susceptible being attacked and the chance that the other would be attacked given that the first had already become infected. This principle can be extended to families of any size, though it has the appreciable disadvantage that it is not clearly related to the epidemiological process known to be taking place. A further discussion of these problems was undertaken by Greenwood (1949). He made the important observation that the failure of the binomial distribution to represent the distribution of cases at stages in the chain where a goodness-of-fit is possible could be due to variations between households in the average chance of infection. Such a state of affairs might well be expected to result from natural variations in both heredity and environment. This remark of Greenwood's is taken as the starting point for the present paper, in which it is shown that a satisfactory mathematical model can be obtained, at least for the small amount of suitable data available, by retaining the hypothesis of a chain of binomial distributions, together with the additional assumption that the average chance of infection varies between different households according to a suitable β-distribution. This involves fitting two parameters instead of only one as previously. Another important modification of detail in Greenwood's original scheme has lately been suggested by Lidwell & Sommerville (1951). Greenwood (1931) assumed that the chance of any susceptible being infected was independent of the number of infectious persons to whom he is exposed. Lidwell & Sommerville, however, considered the alternative assumption that the risk of infection for any susceptible is the same with respect to each infectious individual. This of course affects only families with more than two susceptibles in addition to the primary case. In their analysis of the distributions of the total number of cases of the common cold in households, Lidwell & Sommerville showed that on the whole the Greenwood model was inadequate whereas their own modified scheme fitted quite well. For a useful discussion of the epidemiological aspects of the problem of estimating infectiousness Hope Simpson's (1952) recent paper may be consulted.

Households with only two individuals, i.e. containing only one susceptible in addition to the primary case, permit one to estimate the average chance of cross-infection, but are too small to exhibit a chain of cases. We shall therefore not consider these any further, but shall proceed to a discussion of households of three and four persons, for which there is observational material available for testing. The treatment of larger households is more complicated though similar in principle.

2. ANALYSIS OF HOUSEHOLDS WITH THREE INDIVIDUALS

We will now consider households of three individuals, containing two susceptibles besides the primary case. Let us suppose that the probability of a susceptible contracting the disease on exposure to an infectious case is $p = 1-q$. Then the first generation of secondary cases, following the primary case, will take the values 0, 1 or 2, with frequencies q^2, $2pq$ and p^2. For the values 0 and 2 the epidemic within the household comes to an end; in the first case there is no further source of infection, and in the second there are no more susceptibles. On the other hand, a single secondary case may be followed in the second generation by 0 or 1 tertiary case, with frequencies q and p. The expected and observed numbers of households falling into the four classes are shown in Table 1, which also includes some suitable material taken from the Providence measles experience already referred to (Wilson *et al.* 1939), together with an obvious notation for the different kinds of chain. Wilson *et al.* analysed their material in several different ways, both including and excluding children under 7 months or over 10 years old. Moreover, some of their tables are really relevant to families containing more than three children. We have decided for the present purpose to confine ourselves to the data on households of three, containing just two further susceptibles, apart from the primary case, at risk between the ages of 7 months and 10 years. This seemed to be the most reasonable way of selecting what might be expected to be a fairly homogeneous group of data. In analysing their data Wilson *et al.* estimated the parameter p by equating the expected size of the *total* epidemic to the observed value. This is not necessarily very efficient. They then examined the fit to the four classes shown in Table 1. A more satisfactory procedure is to estimate p by maximum likelihood using all four classes shown in the table. It is easy to see that this efficient estimate is

Table 1. *Greenwood chain for households of three*

Type of chain	Expected no. of households	Observed no. of households	Providence measles data (see text)	Fitted values
$1 \to 0$	nq^2	a	34	14·9
$1 \to 1 \to 0$	$2npq^2$	b	25	23·5
$1 \to 1 \to 1 \to 0$	$2np^2q$	c	36	87·7
$1 \to 2 \to 0$	np^2	d	239	207·9
Total	n	n	334	334·0

$$\hat{p} = (b+2c+2d)/(2a+3b+3c+2d), \tag{1}$$

with large sample variance
$$\operatorname{var}\hat{p} = pq/2n(1+pq). \tag{2}$$

Applying formula (1) to the data shown in Table 1, we find $p = 0{\cdot}789$. The corresponding expectations appear in the right-hand column. The goodness-of-fit χ^2 is 59·8, which for 2 degrees of freedom gives $P < 0{\cdot}001$. Thus the Greenwood model, which for groups of three is unaffected by the modification suggested by Lidwell & Sommerville, is quite inappropriate as it stands. Comparing the actual observations with their expectations in Table 1 shows that there are more families than expected, in which the epidemic either does not go beyond the primary case or else involves the remaining two susceptibles immediately. This fits in with Greenwood's (1949) suggestion that the chance of infection may vary from family to family. Suppose we now assume that this source of variation in p between households is represented by the β-distribution

$$dF = \frac{1}{B(x,y)} p^{x-1}(1-p)^{y-1}\,dp, \quad (0 \leqslant p \leqslant 1). \tag{3}$$

All we have to do is to average the expectations for each kind of chain over all households and exhibit them as functions of x and y. As the individual frequencies in Greenwood or Lidwell chains can always be represented in terms of quantities like $p^r q^s$, it is convenient to work out the expectation of this quantity. We have

$$\begin{aligned} Ep^rq^s &= \int_0^1 p^{x+r-1}q^{y+s-1}\,dp/B(x,y) \\ &= B(x+r,y+s)/B(x,y). \end{aligned}$$

Therefore

$$Ep^rq^s = x(x+1)\dots(x+r-1)\,y(y+1)\dots(y+s-1)/(x+y)(x+y+1)\dots(x+y+r+s-1). \quad (4)$$

In particular

$$Ep^r = x(x+1)\dots(x+r-1)/(x+y)(x+y+1)\dots(x+y+r-1), \quad (5)$$

and

$$Eq^s = y(y+1)\dots(y+s-1)/(x+y)(x+y+1)\dots(x+y+s-1). \quad (6)$$

If we apply these results to the simple expectations shown in Table 1 we obtain the derived set of values set out in Table 2, which we shall proceed to analyse by the familiar technique of maximum-likelihood scoring. The maximum-likelihood scores for x and y are clearly

$$\left.\begin{aligned} S_x &= \left(\frac{b+c+d}{x}+\frac{c+d}{x+1}\right)-\left(\frac{n}{x+y}+\frac{n}{x+y+1}+\frac{b+c}{x+y+2}\right), \\ \text{and} \quad S_y &= \left(\frac{a+b+c}{y}+\frac{a+b}{y+1}\right)-\left(\frac{n}{x+y}+\frac{n}{x+y+1}+\frac{b+c}{x+y+2}\right). \end{aligned}\right\} \quad (7)$$

Table 2. *Modified chain for household of three with variable p*

Type of chain	Expected no. of households	Observed no. of households	Providence measles data (see text)	Fitted values
$1\to0$	$ny(y+1)/(x+y)(x+y+1)$	a	34	34·9
$1\to1\to0$	$2nxy(y+1)/(x+y)(x+y+1)(x+y+2)$	b	25	22·7
$1\to1\to1\to0$	$2nyx(x+1)/(x+y)(x+y+1)(x+y+2)$	c	36	37·6
$1\to2\to0$	$nx(x+1)/(x+y)(x+y+1)$	d	239	238·8
Total	n	n	334	334·0

The expected amounts of information are

$$\left.\begin{aligned} I_{xx} &= \frac{n(x+2y+1)}{x(x+y)(x+y+1)}+\frac{nx(x+3y+2)}{(x+1)(x+y)(x+y+1)(x+y+2)}-J, \\ I_{xy} &= \qquad -J, \\ I_{yy} &= \frac{n(2x+y+1)}{y(x+y)(x+y+1)}+\frac{ny(3x+y+2)}{(y+1)(x+y)(x+y+1)(x+y+2)}-J, \\ \text{where} \quad J &= \frac{n}{(x+y)^2}+\frac{n}{(x+y+1)^2}+\frac{2nxy}{(x+y)(x+y+1)(x+y+2)^2}. \end{aligned}\right\} \quad (8)$$

In many cases it will be easier to use the observed amounts of information, which can be derived from the three different expressions in brackets making up the two scores in (7) simply by replacing the denominators by their squares.

An easy way to obtain two rough estimates of x and y to commence the iterative process is to amalgamate the second and third classes in Table 2, and then equate the observations to their expectations. This gives

$$\left.\begin{aligned} \check{x} &= (b+c)(b+c+2d)/\{4ad-(b+c)^2\}, \\ \text{and} \quad \check{y} &= (b+c)(2a+b+c)/\{4ad-(b+c)^2\}. \end{aligned}\right\} \quad (9)$$

If we now use these results on the Providence measles data, we obtain first the initial estimates from equation (9), giving

$$\check{x} = 1{\cdot}142 \quad \text{and} \quad \check{y} = 0{\cdot}273,$$

for which values the scores are $S_x = -1{\cdot}1253$ and $S_y = +2{\cdot}1255$. The information and covariance matrices are respectively

$$\{I\} = \begin{bmatrix} 63{\cdot}8027 & -226{\cdot}696 \\ -226{\cdot}696 & 1075{\cdot}635 \end{bmatrix} \quad \text{and} \quad \{V\} = 10^{-2} \times \begin{bmatrix} 6{\cdot}2401 & +1{\cdot}3151 \\ +1{\cdot}3151 & 0{\cdot}3701 \end{bmatrix}. \tag{10}$$

The adjustments to be made to $\check{x}$ and $\check{y}$ are therefore $-0{\cdot}042$ and $-0{\cdot}007$ giving 1·100 and 0·266, respectively. A further stage of iteration makes a slight change to

$$\hat{x}_3 = 1{\cdot}092 \pm 0{\cdot}250 \quad \text{and} \quad \hat{y}_3 = 0{\cdot}264 \pm 0{\cdot}061, \tag{11}$$

where we have followed the estimates by their standard errors derived from the covariance matrix, and the subscripts refer to the size of the household. The expected values are given in the right-hand column of Table 2. The fit to the data is now very good indeed, the χ^2 with one degree of freedom being about 0·32. It is interesting to observe that since y is less than unity the β-distribution is J-shaped with an infinite ordinate at $p = 1$. This is in keeping with the excess of families observed in Table 1, in which both susceptibles succumbed immediately to infection by the primary case.

3. ANALYSIS OF HOUSEHOLDS WITH FOUR INDIVIDUALS

When we come on to households with a total of four susceptibles, i.e. three in addition to the primary case, then the modification introduced by Lidwell & Sommerville begins to have an effect. It influences the tertiary distribution, in which, when there are exactly two secondary cases, the values 0 and 1 will occur with probabilities q^2 and $1-q^2$ instead of q and p. Let us again use the Providence measles data, confining ourselves to households of four, containing just three further susceptibles, apart from the primary cases, at risk between the ages of 7 months and 10 years. The appropriate expectations and observations are all collected together in Table 3. The maximum likelihood estimate for the Greenwood model is plainly

$$\hat{p} = (b + 2c + 2d + 3e + 3f + 3g + 3h)/(3a + 5b + 6c + 4d + 6e + 5f + 4g + 3h). \tag{12}$$

Table 3. *Greenwood and Lidwell chains for households of four*

Type of chain	Expected no. of households		Observed no. of households	Providence data	Fitted values	
	Greenwood	Lidwell			Greenwood	Lidwell
$1 \to 0$	nq^3	nq^3	a	4	0·9	1·2
$1 \to 1 \to 0$	$3npq^4$	$3npq^4$	b	3	0·4	0·7
$1 \to 1 \to 1 \to 0$	$6np^2q^4$	$6np^2q^4$	c	1	0·7	1·0
$1 \to 2 \to 0$	$3np^2q^2$	$3np^2q^3$	d	8	8·2	2·2
$1 \to 1 \to 1 \to 1 \to 0$	$6np^3q^3$	$6np^3q^3$	e	4	2·7	3·4
$1 \to 1 \to 2 \to 0$	$3np^3q^2$	$3np^3q^2$	f	3	6·5	7·3
$1 \to 2 \to 1 \to 0$	$3np^3q$	$3np^2q(1-q^2)$	g	10	31·0	38·7
$1 \to 3 \to 0$	np^3	np^3	h	67	49·6	45·5
Total	n	n	n	100	100·0	100·0

Substituting the Providence data gives $p = 0{\cdot}791$, and the expected numbers are as shown in the table. The goodness-of-fit test, amalgamating the first four classes and then the next two, gives $P < 0{\cdot}001$. To estimate p on the Lidwell model we have to solve

$$S_p = \frac{b+2c+2d+3e+3f+2g+3h}{p} - \frac{3a+4b+4c+3d+3e+2f+g}{q} + \frac{2qg}{1-q^2} = 0. \qquad (13)$$

Taking a few trial values and then interpolating inversely soon gives the required root. We find, for the data of Table 3, that $p = 0{\cdot}769$. Again a very poor fit is obtained with $P < 0{\cdot}001$.

Table 4. *Modified Greenwood chain for households of four with variable* p

Type of chain	Expected no. of households	Observed no. of households	Providence data	Fitted values
$1\to 0$	$ny(y+1)(y+2)/(x+y)\ldots(x+y+2)$	a	4	4·9
$1\to 1\to 0$	$3nxy(y+1)(y+2)(y+3)/(x+y)\ldots(x+y+4)$	b	3	2·6
$1\to 1\to 1\to 0$	$6nx(x+1)y(y+1)(y+2)(y+3)/(x+y)\ldots(x+y+5)$	c	1	1·9
$1\to 2\to 0$	$3nx(x+1)y(y+1)/(x+y)\ldots(x+y+3)$	d	8	5·0
$1\to 1\to 1\to 1\to 0$	$6nx(x+1)(x+2)y(y+1)(y+2)/(x+y)\ldots(x+y+5)$	e	4	2·1
$1\to 1\to 2\to 0$	$3nx(x+1)(x+2)y(y+1)/(x+y)\ldots(x+y+4)$	f	3	3·0
$1\to 2\to 1\to 0$	$3nx(x+1)(x+2)y/(x+y)\ldots(x+y+3)$	g	10	13·3
$1\to 3\to 0$	$nx(x+1)(x+2)/(x+y)\ldots(x+y+2)$	h	67	67·1
Total	n	n	100	99·9

We now turn to our previous assumption that p varies between households and has the β-distribution given by (3). Applying this modification to the expectations of the Greenwood model in a manner similar to that employed for households of size three, we obtain the expectations and observations shown in Table 4, which is analogous to Table 2. The maximum likelihood scores are

$$\left.\begin{aligned}
S_x &= \left(\frac{n-a}{x} + \frac{n-a-b}{x+1} + \frac{e+f+g+h}{x+2}\right)\\
&\qquad - \left(\frac{n}{x+y} + \frac{n}{x+y+1} + \frac{n}{x+y+2} + \frac{n-a-h}{x+y+3} + \frac{b+c+e+f}{x+y+4} + \frac{c+e}{x+y+5}\right),\\
S_y &= \left(\frac{n-h}{y} + \frac{n-g-h}{y+1} + \frac{a+b+c+e}{y+2} + \frac{b+c}{y+3}\right) - \left(\ldots\right),
\end{aligned}\right\} \qquad (14)$$

where the second member of S_y in brackets is the same as the second member of S_x. The information functions are

$$\left.\begin{aligned}
I_{xx} &= \left(\frac{n-a}{x^2} + \frac{n-a-b}{(x+1)^2} + \frac{e+f+g+h}{(x+2)^2}\right) - K,\\
I_{xy} &= \qquad\qquad\qquad\qquad\qquad\qquad\quad -K,\\
I_{yy} &= \left(\frac{n-h}{y^2} + \frac{n-g-h}{(y+1)^2} + \frac{a+b+c+e}{(y+2)^2} + \frac{b+c}{(y+3)^2}\right) - K,\\
&\text{where}\\
K &= \frac{n}{(x+y)^2} + \frac{n}{(x+y+1)^2} + \frac{n}{(x+y+2)^2} + \frac{n-a-h}{(x+y+3)^2} + \frac{b+c+e+f}{(x+y+4)^2} + \frac{c+e}{(x+y+5)^2}.
\end{aligned}\right\} \qquad (15)$$

Some additional accuracy can no doubt be gained if the observed numbers in the several numerators in (15) are replaced by their expectations. There seems to be no obvious way of finding simple initial estimates, but with a few trial values it is not difficult to get sufficiently accurate approximations. Scoring with $x = 1{\cdot}43$ and $y = 0{\cdot}31$ gives

$$S_x = +0{\cdot}4124 \quad \text{and} \quad S_y = -0{\cdot}2782.$$

The information and covariance matrices are

$$\{I\} = \begin{bmatrix} 14{\cdot}6024 & -55{\cdot}233 \\ -55{\cdot}233 & 303{\cdot}810 \end{bmatrix} \quad \text{and} \quad \{V\} = 10^{-2} \times \begin{bmatrix} 21{\cdot}925 & +3{\cdot}9861 \\ +3{\cdot}9861 & 1{\cdot}0538 \end{bmatrix}. \tag{16}$$

The required adjustments to the estimates are therefore $+0{\cdot}08$ and $+0{\cdot}01$. We may therefore write the maximum-likelihood estimates as

$$\hat{x}_4 = 1{\cdot}51 \pm 0{\cdot}47 \quad \text{and} \quad \hat{y}_4 = 0{\cdot}32 \pm 0{\cdot}10. \tag{17}$$

Again, the expected values appearing in Table 4 show a very close fit to the observations. Grouping the first six classes into three pairs leaves 4 degrees of freedom altogether. The resulting goodness-of-fit χ^2 with 2 degrees of freedom is 2·20, a value which is entirely satisfactory. It hardly seems worth while at present to undertake a similarly modified treatment of the Lidwell chain model.

Comparing the results given in (11) with those in (17) shows no significant difference between the estimates obtained for households of three and households of four. We may therefore combine these estimates. The weighted averages, with weights equal to the reciprocals of the variances, are

$$\hat{x} = 1{\cdot}18 \pm 0{\cdot}22 \quad \text{and} \quad \hat{y} = 0{\cdot}279 \pm 0{\cdot}052. \tag{18}$$

The average chance of infection, $\bar{p}$, is given by putting $r = 1$ in (5). Thus we obtain

$$\bar{p} = x/(x+y) = 0{\cdot}81, \tag{19}$$

using the values appearing in (18). If in the future it proves possible to group data on the household distribution of the cases of a disease according to some such classifications as geographical area or social class, etc., it will be interesting to see whether p has a more circumscribed distribution within each of these groups.

4. SUMMARY

Greenwood (1931) was the first to introduce the chain-binomial model to describe the household distribution of cases of a disease with a short period of high infectivity and a constant incubation period. Greenwood obtained satisfactory goodness-of-fits to the St Pancras measles data, at least so far as the distribution of the total number of cases was concerned. Wilson *et al.* (1939) showed that on their measles data for Providence, Rhode Island, even if the fit to the distribution of the total number of cases was satisfactory, the theory was quite inadequate when the distributions of the separate generations of the disease within a household were examined. Later, Greenwood (1949) suggested that this might be due to variation in the chance of infection, p, between households. In the present paper the Providence data are re-examined on the assumption that p varies between households according to a suitable β-distribution. Two parameters then have to be estimated. p appears

to have a J-shaped distribution, with an infinite ordinate at $p = 1$. Two groups of the Providence data for households of three and four individuals were investigated, and in both cases excellent fits of the theory to the data were obtained.

REFERENCES

Bailey, N. T. J. (1953). The total size of a general stochastic epidemic. *Biometrika*, **40**, 177.

Greenwood, M. (1931). On the statistical measure of infectiousness. *J. Hyg., Camb.*, **31**, 336.

Greenwood, M. (1949). The infectiousness of measles. *Biometrika*, **36**, 1.

Hope Simpson, R. E. (1952). Infectiousness of communicable diseases in the household. *Lancet*, **2**, 549.

Lidwell, O. M. & Sommerville, T. (1951). Observations on the incidence and distribution of the common cold in a rural community during 1948 and 1949. *J. Hyg., Camb.*, **49**, 365.

Wilson, E. B., Bennett, C., Allen, M. & Worcester, J. (1939). Measles and scarlet fever in Providence, R.I., 1929–34 with respect to age and size of family. *Proc. Amer. Phil. Soc.* **80**, 357.

Wilson, E. B. (1947). The spread of measles in the family. *Proc. Nat. Acad. Sci., Wash.*, **33**, 68.

SPREAD OF DISEASES IN A RECTANGULAR PLANTATION WITH VACANCIES

By G. H. FREEMAN

East Malling Research Station

1. Introduction

This paper discusses the spread of diseases from plant to plant in a rectangular plantation where there may be some missing plants, which will be described as 'vacancies'. The problem is considered in two ways: (i) the degree of agglomeration of diseased plants on any one occasion, (ii) the spread of disease over a period of time. In the work which gave rise to this investigation the plants were hops, and the spread of the virus disease, nettlehead, over a period of five years was examined.

The methods used for the first aspect of the investigation, which may conveniently be called *spatial*, derive from various writers, principally Krishna Iyer. Historically, the first attempt to deal with this problem was that of Todd (1940), although previously Cochran (1936) had examined the linear case. Todd considered the distribution of pairs of adjacent plants, or points on a lattice, on the assumption that the underlying distribution was the binomial. The method of *doublets*, as he called it, has been followed in most of the subsequent work, although Finney (1947) showed empirically that Todd's formula seriously over-estimated the variance, because in an agglomeration doublets do not occur independently of one another. Van der Plank (1946) applied Cochran's and Todd's methods to virological studies, but without knowing of Finney's criticisms. Furthermore, he only considered doublets in one direction.

Moran (1948) considered the distribution of doublets for a general lattice, not specifically rectangular, and also found third and fourth moments in a particular case of the rectangular lattice. Krishna Iyer gave general solutions to the problem for a rectangular lattice in two papers; diagonal pairs were excluded in the 1950 paper, but considered in that of 1949.

Moran and Krishna Iyer both distinguished two cases: (i) where the probability of any plant being diseased is the same as that of any other plant being diseased and independent of the number of diseased plants, and (ii) where the number of diseased plants is known and interest attaches only to their distribution over the area. These cases are called *free* and *non-free* sampling respectively. Following the accepted notation, diseased plants are referred to here as *black* and healthy plants as *white*, and it is the distributions of both black-black (BB) joins and black-white (BW) joins that are of interest in practice.

Todd also considered sets of three and four neighbouring plants, called respectively *triplets* and *quadruplets*, but for these Finney (1947) showed that the variance would be under-estimated on the binomial assumption. Sukhatme (1949) considered triplets and stated that the inclusion of diagonals gave a more efficient test than the exclusion of diagonals, i.e. resulted in smaller coefficients of variation. He showed, further, that when diagonals were included the coefficients of variation for triplets were larger than those for doublets. In consequence of these results no further attention is here paid to triplets.

None of the above-mentioned writers has considered what happens when there are vacancies on the lattice, although Moran's general formula can be adapted to this case. In the present paper the approach to the problem has been to generalize Krishna Iyer's methods for a rectangular lattice to make them applicable where there are vacancies rather than to particularize Moran's methods. The results must of course be the same, but Krishna Iyer's methods are the more elegant, for calculating both the probabilities of the number of joins of a given type and the total number of joins. (For this latter, the method of his 1949 paper has been preferred.) Both the cases, including diagonals and excluding diagonals, are considered, these being referred to, by analogy with chess, as *queen* and *rook* cases respectively. In the particular example given later in the paper the queen case was thought appropriate, but formulae are given here for the rook case as well, for completeness.

The methods used for the second aspect of the investigation, which may be called *temporal*, derive from the work of Barnard (1947) and Pearson (1947). Barnard distinguished three possible types of significance test for a 2 × 2 table—of which the second or 2 × 2 comparative trial is appropriate here—and Pearson obtained formulae for actually testing the significance of observed results. All that is done here on this aspect is to consider the relevant method of analysis and apply it.

2. SPATIAL ASPECT

Consider a rectangular lattice m by n with vacancies, where $m+n = a$, $m \times n = b$. Suppose that there are u vacancies in the interior of the lattice, v on the border and w at the corners, where $u+v+w = U$. Points that are not vacancies will be called *existent* points. In the free sampling case let the probabilities of getting black and white points be respectively p and q, and in the non-free sampling case let the numbers of black and white points be respectively n_1 and n_2. These two colours need not necessarily account for all the existent points in either case. Only the non-free sampling case will be considered in detail, this being slightly the more complicated and also the more likely to be of practical use. The appropriate substitutions for free sampling will be indicated. All the probabilities derived below assume the points to have been selected at random and their colours then examined.

In the non-free sampling case, then, the probability of any point being black is $\frac{n_1}{b-U}$ and of any point being white is $\frac{n_2}{b-U}$. Thus the probability of any join of two points being BB is

$$\frac{n_1(n_1-1)}{(b-U)(b-U-1)} \quad \text{or} \quad \frac{n_1!(b-U-2)!}{(n_1-2)!(b-U)!}. \tag{1}$$

Similarly the probability of a BW join is

$$\frac{2n_1 n_2}{(b-U)(b-U-1)}. \tag{2}$$

The corresponding probabilities for free sampling are p^2 and $2pq$.

Further, two joins can arise in only two ways. Either they can come from three adjacent points, in which case they will be described as *dependent* pairs of joins, or they can come from four points in two discrete pairs, when they will be called *independent* pairs of joins. For non-free sampling the probability of obtaining three adjacent black points is

$$\frac{n_1!(b-U-3)!}{(n_1-3)!(b-U)!}, \tag{3}$$

and of obtaining four black points in two discrete pairs is

$$\frac{n_1!(b-U-4)!}{(n_1-4)!(b-U)!}. \tag{4}$$

For BW joins the position is slightly different. If three adjacent points are to make two dependent BW pairs the points must be either BWB or WBW. The sum of the probabilities of the different kinds of dependent BW pairs is thus

$$\frac{n_1(n_1-1)\,n_2+n_1n_2(n_2-1)}{(b-U)\,(b-U-1)\,(b-U-2)} = \frac{n_1n_2(n_1+n_2-2)}{(b-U)\,(b-U-1)\,(b-U-2)}. \tag{5}$$

The probability of two independent BW pairs is

$$\frac{4n_1!\,n_2!\,(b-U-4)!}{(n_1-2)!\,(n_2-2)!\,(b-U)!}. \tag{6}$$

The probabilities corresponding to (3)–(6) for free sampling are respectively p^3, p^4, $pq(p+q)$, $4p^2q^2$.

Now Krishna Iyer (1950) has shown that the rth factorial moment about zero for the distributions of both BB and BW joins is $r!$ times the sum of the expectations of the different ways of obtaining r joins of that type in the lattice. The argument employed in that paper (pp. 199–201) can be readily modified to show that the above statement holds for both rook and queen cases, even if there are vacancies.

All that is necessary, therefore, is to calculate the numbers of joins or sets of joins of a given type for the particular case under examination, and this is done below.

(a) *Queen case.* On the lattice there are, in all, 4 corner points, $2(a-4)$ border points and $b-2a+4$ interior points. Thus the total number of joins of existent points on the lattice is A, where
$$2A = 3\,.\,4+5\,.\,2(a-4)+8(b-2a+4)-2V,$$
V being a correction for vacancies, i.e.
$$A = 4b-3a+2-V. \tag{7}$$

Now each interior vacancy contributes 8 to V, each border vacancy 5 and each corner vacancy 3, except that two neighbouring vacancies reduce V by 1. Thus, if p is the number of pairs of vacancies
$$V = 8u+5v+3w-p. \tag{8}$$

If μ is the expected number of joins of a given type we have, for non-free sampling from the appropriate expression (1) or (2): for BB joins

$$\mu = \frac{An_1!\,(b-U-2)!}{(n_1-2)!\,(b-U)!}, \tag{9}$$

and for BW joins
$$\mu = \frac{2An_1n_2}{(b-U)\,(b-U-1)}, \tag{10}$$

where A is derived from (7) and (8). For free sampling p^2 and $2pq$ are substituted for (1) and (2) respectively.

The numbers of dependent and independent pairs of joins must now be considered. In the first case, where there are three adjacent points, two joins can arise in B ways, where
$$B = 3\,.\,4+10\,.\,2(a-4)+28(b-2a+4)-W,$$
W being a correction for vacancies, i.e.
$$B = 4(7b-9a+11)-W. \tag{11}$$

Now each interior vacancy contributes 28 to W directly, each border vacancy 10 and each corner vacancy 3. Also each interior existent point with one neighbouring vacancy contributes a further 7 to W, one with two neighbouring vacancies $7+6\ (=13)$, one with three $7+6+5$, etc. Similarly, border existent points with one neighbouring vacancy contribute 4 to W, those with two $4+3$, etc. Finally, corner existent points with one neighbouring vacancy contribute 2 to W, etc. Let there be x_i interior existent points with i neighbouring vacancies, y_i border existent points with i neighbouring vacancies and z_i corner existent points with i neighbouring vacancies. Then

$$W = 28u+10v+3w+(7x_1+13x_2+18x_3+22x_4+25x_5+27x_6+28x_7+28x_8) \\ +(4y_1+7y_2+9y_3+10y_4+10y_5)+(2z_1+3z_2+3z_3). \qquad (12)$$

Since the total number of pairs of joins is $\frac{A(A-1)}{2}$, and pairs can only be dependent or independent, the number of independent pairs is $\frac{A(A-1)}{2}-B$.

For BB joins we thus have, from (3) and (4), that the second factorial moment is $\mu'_{[2]}$, where

$$\mu'_{[2]} = \frac{2Bn_1!\,(b-U-3)!}{(n_1-3)!\,(b-U)!}+\frac{[A(A-1)-2B]\,n_1!\,(b-U-4)!}{(n_1-4)!\,(b-U)!}.$$

Now if σ^2 is the variance, $\qquad \sigma^2 = \mu'_{[2]}+\mu-\mu^2. \qquad (13)$

Hence
$$\sigma^2 = \frac{2Bn_1!\,(b-U-3)!}{(n_1-3)!\,(b-U)!}+\frac{An_1!\,(b-U-2)!}{(n_1-2)!\,(b-U)!} \\ +\frac{[A(A-1)-2B]\,n_1!\,(b-U-4)!}{(n_1-4)!\,(b-U)!}-\left[\frac{An_1!\,(b-U-2)!}{(n_1-2)!\,(b-U)!}\right]^2. \qquad (14)$$

Similarly, we have for BW joins, by substituting in (13) from (5), (6) and (10),

$$\sigma^2 = \frac{2Bn_1n_2(n_1+n_2-2)}{(b-U)(b-U-1)(b-U-2)}+\frac{2An_1n_2}{(b-U)(b-U-1)} \\ +\frac{4[A(A-1)-2B]\,n_1!\,n_2!\,(b-U-4)!}{(n_1-2)!\,(n_2-2)!\,(b-U)!}-\left[\frac{2An_1n_2}{(b-U)(b-U-1)}\right]^2. \qquad (15)$$

The only change necessary for the free sampling case is the substitution of the appropriate probability expressions in the first and second factorial moments. For both free and non-free sampling Krishna Iyer has shown that the distributions tend to the normal form as the size of the lattice increases, if there are no vacancies. Thus, if the number of vacancies remains finite these distributions here will also tend to the normal in the limiting case. This will be true for both queen and rook cases.

(*b*) *Rook case.* The number of points on the lattice being the same as before, the total number of joins of existent points on the lattice is A', where

$$2A' = 2\,.\,4+3\,.\,2(a-4)+4(b-2a+4)-2V',$$

V' being a correction for vacancies, i.e.

$$A' = 2b-a-V'. \qquad (16)$$

Now each interior vacancy contributes 4 to V', each border vacancy 3 and each corner vacancy 2, except that two neighbouring vacancies reduce V' by 1. (Here, and for the remainder of the section devoted to the rook case, two points are only regarded as neighbours if they are next to one another on the rank or file.) Thus, if p' is the number of pairs of vacancies, p' being in general different from p as defined above, we have

$$V' = 4u+3v+2w-p'. \tag{17}$$

The expressions (9) and (10) in the queen case are now replaced by the following values for the mean number of joins: for BB joins

$$\mu = \frac{A'n_1!(b-U-2)!}{(n_1-2)!(b-U)!}, \tag{18}$$

and for BW joins

$$\mu = \frac{2A'n_1n_2}{(b-U)(b-U-1)}, \tag{19}$$

A' being derived from (16) and (17). For free sampling p^2 and $2pq$ are again substituted for (1) and (2) respectively.

For the calculation of the second moment we again need to consider the number of ways of getting the two different types of pairs of joins. Now, three adjacent points can here be derived in B' ways, where

$$B' = 1.4+3.2(a-4)+6(b-2a+4)-W',$$

W' being a correction for vacancies, i.e.

$$B' = 6b-6a+4-W'. \tag{20}$$

Now each interior vacancy contributes 6 to W' directly, each border vacancy 3 and each corner vacancy 2. Also each interior existent point with one neighbouring vacancy contributes a further 3 to W', one with two neighbouring vacancies $3+2$, etc. Similarly, border existent points with one neighbouring vacancy contribute 2, and so on, and corner existent points with one or two neighbouring vacancies 1. Let there be x'_i interior existent points with i neighbouring vacancies, y'_i border existent points with i neighbouring vacancies and z'_i corner existent points with i neighbouring vacancies, where x'_i is not in general the same as x_i, etc. Thus

$$W' = 6u+3v+2w+(3x'_1+5x'_2+6x'_3+6x'_4)+(2y'_1+3y'_2+3y'_3)+(z'_1+z'_2). \tag{21}$$

Since the total number of pairs of joins is $\dfrac{A'(A'-1)}{2}$, the number of independent pairs is, by the same reasoning as for the queen case, $\dfrac{A'(A'-1)}{2}-B'$.

Hence for BB joins we have, by substituting in (13) from (3), (4) and (9), the same expression as (14) with A' written for A and B' for B. Similarly, the variance of the number of BW joins is given by making the necessary modifications in (15). As before, the changes required for the free sampling case are merely in the probability expressions.

3. APPLICATIONS

The method as outlined above is completely general, but its application can best be described by means of a specific example, for further details of which see the paper of Legg (1953). The apparent spread of hop nettlehead disease was considered over a period of five years, to determine whether it could be regarded as being due to a natural transport of inoculum from plant to plant. In order to do this it is essential to examine both spatial and temporal aspects of the spread.

The spatial approach is necessary for the consideration of whether diseased plants occur in agglomerations, or whether they can be regarded as occurring randomly in space. The hops in this instance were planted at a distance of 6 ft. square, and the method of training them was such that each plant touched its neighbours, along the rows, across the rows and diagonally. This meant that the queen case suggested itself as being the one to use. The result of the analysis was that there was strong evidence, in three out of the five years at least, that diseased plants were more often found in agglomerations than would have been expected had these plants been randomly distributed.

As an example of the use of the method Legg's data for the year 1951 will be considered. Fig. 1 shows the distribution of plants showing nettlehead symptoms (black plants), those without symptoms (white plants), and missing plants (vacancies). The lattice is 11×30 plants in area, and so $a = 41$, $b = 330$. Further, there are 122 plants with symptoms and 196 without symptoms, and so these are the respective values of n_1 and n_2.

The distribution of vacancies is such that $u = 11$, $v = 1$, $w = 0$, and there are no pairs of neighbouring vacancies, i.e. $p = 0$. We thus have

$$U = 11+1 = 12 \quad \text{and} \quad V = (8\times 11)+(5\times 1) = 93.$$

This gives $b-U = 318$, and, from equation (7),

$$A = (4\times 330)-(3\times 41)+2-93 = 1106.$$

We now have to consider the existent points with various numbers of neighbouring vacancies, and we find that $x_1 = 66$, $x_2 = 8$, $x_3 = 1$, $y_1 = 8$, while the remaining x's and y's, and all the z's, are zero. Thus, from equation (12),

$$W = (28\times 11)+(10\times 1)+[(7\times 66)+(13\times 8)+(18\times 1)]+(4\times 8) = 934,$$

and so, from equation (11)

$$B = 4[(7\times 330)-(9\times 41)+11]-934 = 6874.$$

Before the next stage of the computation it is convenient to find $A(A-1)-2B$, which equals $(1106\times 1105)-6874 = 1208382$. For BB joins we then have

$$\mu = \frac{1106\times 122\times 121}{318\times 317} = 162{\cdot}0$$

and $$\sigma^2 = \mu + \frac{2\times 6874\times 122\times 121\times 120}{318\times 317\times 316} + \frac{1208382\times 122\times 121\times 120\times 119}{318\times 317\times 316\times 315} - \mu^2,$$

which reduces to $530{\cdot}66$, whence $\sigma = 23{\cdot}0$.

Similarly, for BW joins, we find $\mu = 524{\cdot}7$, $\sigma = 15{\cdot}8$.

The observed numbers of BB and BW joins are respectively 220 and 408, the former of which exceeds expectation by about two and a half times the standard error, and the latter of which falls short of expectation by more than seven times the standard error. By thus using the standard error as a measure of the discrepancy between theory and observation it appears from the combination of these two results that it is very unlikely that the observed distribution of diseased plants was in fact due to a random distribution of these plants in the field.

In the case of nettlehead, which is a virus disease probably transmitted by insects, and with this particular method of growing hops, the queen case of the spatial aspect must be the one to use. This need not, however, always be the case, and there may be times when it is desirable to consider the spread of disease where diagonal joins are excluded. Furthermore, in the case of a disease believed to be spread by cultivations, these being carried out in one direction only, it may be appropriate to consider a linear lattice. (Since the formula for this case can be easily derived from that for a rectangular lattice by taking one of the dimensions to be of unit width, it is not given explicitly here.) There may, moreover, be cases where practically nothing is known about the disease under investigation, and then it may be useful to consider a sort of sequential approach, first using a linear lattice, and then, successively, the rook and queen cases for the rectangular lattice. If this is done, and one method shows evidence of spread but the next does not, the results may well shed some light on the method of spread of the disease in question, and thus provide clues to the agency of spread.

Fig. 1. Nettlehead status of plants in 1951

	39	40	41	42	43	44	45	46	47	48	49
1	.	.	.	.	.	N	V	N	.	N	N
2	.	N	.	.	N	N	N	.	N	.	.
3	N	N	N	N	.	N	.	N	N	N	.
4	N	N	.	N	N	N	N	N	N	.	.
5	N	N	N	N	.	N	.	.	N	N	.
6	.	N	N	.	N	N	N	.	.	N	N
7	N	N	N	N	.	N	.	.	.	.	.
8	N	.	V	N	.	N	N	N	N	N	N
9	N	N	.	N	V	.	N	N	.	N	.
10	.	N	N	N	N	N	N	N	V	.	N
11	N	N	N	.	.	.	N	N	.	.	N
12	.	.	N	.	.	N	.	N	N	N	N
13	N	N	V	N	.	N	.	.	N	.	N
14	N	.	N	N	.	N	V	.	V	.	.
15	.	V	.	.	.	.	.	.	.	N	.
16	.	.	.	.	N	.	V	.	N	N	N
17	.	.	.	.	.	.	.	.	.	V	N
18	.	.	N	.	N	.	.	.	.	N	.
19	N	.	.	N	.	.	.	.	.	.	.
20	.	.	.	N	.	.	.	.	N	.	.
21	.	.	V	.	.	N	.	.	N	.	.
22	.	.	.	.	.	N	.	N	N	.	N
23	.	.	.	.	N	V	.	.	.	.	N
24	.	.	N	.	.	.	.	.	.	.	N
25	.	.	.	.	.	.	.	N	.	N	N
26	.	.	.	.	.	.	.	.	N	N	N
27	.	.	.	.	.	.	.	.	.	N	N
28	N	.	.	.	.	.	.	.	.	N	.
29	.	.	.	.	.	.	.	.	.	N	.
30	.	.	.	.	.	.	.	.	.	.	.

N = Nettlehead plant. V = Vacancy. . = Plant without symptoms.

Investigation of the spatial aspect of the matter is not in general enough by itself, and the temporal approach should be used in conjunction with it. The reason for this is that even definite evidence from the spatial approach cannot do more than suggest that diseased plants occur in clusters. This is not sufficient to prove that there is infectious spread, for there may be ecological or other factors which could give rise to the same pattern of diseased plants. For example, there could be localized conditions of soil or shelter giving rise to patches where the disease develops or where symptoms are expressed. Equally a physiological disorder can occur in patches. If, however, the temporal method shows that the disease is spreading throughout the plantation under investigation, while retaining the characteristic pattern of agglomerations, there would be strong evidence of genuinely infectious spread.

The investigation of nettlehead disease provides one practical instance of the use of the two approaches, and a second example is given by another disease of hops, the fluctuating type of *Verticillium* wilt. In one particular locality the appearance in one year of agglomerations of diseased plants suggested a spread of the disease, but these agglomerations were found to be due to the influence of flood water in the winter in relation to the contour pattern in this hop garden. Here the use of the temporal approach would have immediately shown the spread to be apparent rather than real.

The temporal method is used to examine whether plants next to diseased plants are more liable to go down with a disease than those all of whose neighbours are healthy. To revert to the nettlehead example, the investigation concerned itself with the virus status in any year of plants healthy in the preceding year, in order to see if the disease spread principally to those plants whose neighbours were already diseased. Any plant which was missing the second year but not diseased the first year was tacitly excluded.

To investigate this, the method of Barnard's 2 × 2 comparative trial was used. For, as Barnard states, the fundamental subject-matter of a 2 × 2 comparative trial is two populations, and here we have the two populations of healthy plants: (i) those having some diseased neighbours, and (ii) those having all healthy neighbours. These populations being given, records are then taken in the next year to see whether the proportion of diseased plants is the same in the two populations. It will be seen that, whereas there may be some doubt as to which type of spatial approach to use, the essential supplement using the temporal approach must make use of the 2 × 2 comparative trial. It is obvious that the same definition of 'neighbours' will obtain for the two aspects of the same investigation.

Fig. 2 shows the distribution of plants showing nettlehead in 1951 and not in 1950 in relation to plants that showed nettlehead in 1950. The division of the population described above shows the following numbers of plants in each category:

Neighbours in 1950	Plants in 1951 healthy in 1950		
	Diseased	Healthy	Total
Some diseased	53	88	141
All healthy	12	90	102
Total	65	178	243

In Barnard's notation, u being a variate distributed normally with unit variance, we then have

$$u = \frac{(53 \times 90) - (12 \times 88)}{243} \div \sqrt{\frac{141 \times 102 \times 65 \times 178}{243 \times 243 \times 242}},$$

whence $u = 4{\cdot}48$ with probability $P < 0{\cdot}001$.

It thus appears that diseased plants are very much more likely to occur if some of their neighbours were previously diseased than if all their neighbours were previously healthy, which suggests that there is some form of infectious spread taking place.

In the nettlehead example, the results of the investigation of the temporal aspect agreed very closely with those of the spatial aspect because those years when the spatial approach suggested agglomeration were those in which there was a tendency for more newly diseased plants to be next to those diseased the previous year. Also, in this instance, the disease had spread so far across the recorded area by the end of the period under investigation as to suggest that the liability of plants to infection was not related to their position, and thus there was less likelihood of ecological factors accounting for the spread. Further, in general, when there were significantly more *BB* joins than expectation there were significantly fewer *BW* joins, and when one did not differ significantly from expectation neither did the other.

Fig. 2. Nettlehead status of plants in 1951 in relation to that of 1950

	39	40	41	42	43	44	45	46	47	48	49
1	.	.	.	.	.	O	V	O	O	O	O
2	.	N	.	.	N	N	O	.	O	.	.
3	O	O	N	N	O	O	.	N	O	N	.
4	O	O	.	N	O	O	O	N	N	O	.
5	N	N	N	O	O	N	O	.	O	O	O
6	.	O	O	.	N	O	N	.	O	O	O
7	O	N	O	O	.	O	.	.	.	.	O
8	N	.	V	N	O	O	N	O	N	N	N
9	N	O	O	O	V	O	N	O	O	N	.
10	.	N	N	O	O	N	N	O	O	.	O
11	O	N	O	O	.	.	O	N	.	.	N
12	.	.	N	.	.	O	.	N	N	O	N
13	O	N	V	O	.	N	.	.	N	O	N
14	O	.	N	N	.	N	V	.	V	.	.
15	.	V	O	.	O	.	.	.	.	O	.
16	.	.	.	.	O	.	V	.	O	O	O
17	.	.	.	.	O	.	.	.	.	V	N
18	.	.	N	.	N	.	.	.	.	O	.
19	N	.	.	N	.	.	.	.	.	.	.
20	.	.	.	N	.	.	.	.	N	.	.
21	.	.	V	.	.	N	.	.	N	.	.
22	.	.	.	.	.	N	.	N	N	.	N
23	.	.	.	.	O	V	.	.	.	.	N
24	.	.	O	.	.	.	.	.	.	.	O
25	.	.	.	.	.	.	.	N	.	N	O
26	.	.	.	.	O	.	.	.	N	O	O
27	.	.	.	.	.	.	.	.	.	N	N
28	N	.	.	.	.	.	.	.	.	N	.
29	.	.	.	.	.	.	.	.	.	N	.
30	.	.	.	.	.	.	.	.	.	.	.

O = Nettlehead plant in 1950, irrespective of whether it had nettlehead, had been replanted or was a vacancy in 1951.
N = Nettlehead plant in 1951 but not 1950.
V = Vacancy in 1951, not nettlehead in 1950.
. = Plant without symptoms in both 1951 and 1950.

In most practical cases plants will be either diseased or healthy, and so the black and white colours are likely to account for all the existent points on the lattice. There could, however, be cases, such as those of a complex of viruses, where more colours were possible. Also, there could be cases where some of the plants were carriers of the disease without showing symptoms themselves. It would then probably be desirable to treat such plants as vacancies from the spatial point of view but as diseased neighbours the previous year in the temporal approach.

4. SUMMARY

This paper considers the spread of disease in a rectangular plantation where there may be some missing plants, or vacancies. The problem is considered from two points of view, spatial and temporal. In the former, formulae are given which modify the existing methods derived by Krishna Iyer to allow for vacancies. In the latter, the 2 × 2 comparative trial of Barnard is used. Both methods are applied to a particular example of the apparent spread of a hop virus disease, nettlehead. In this case the results achieved by the two methods supplement each other, and, in general, both must be used before it can be concluded if the apparent spread corresponds to a spread of inoculum.

I am much indebted to Dr S. C. Pearce for helpful discussion on the statistical aspects, to Mr J. T. Legg for his ready co-operation on the plant pathology side of the study and to Dr R. V. Harris for bringing to my notice the example relating to *Verticillium* wilt.

REFERENCES

BARNARD, G. A. (1947). Significance tests for 2 × 2 tables. *Biometrika*, **34**, 123–38.

COCHRAN, W. G. (1936). The statistical analysis of field counts of diseased plants. *J.R. Statist. Soc. Suppl.* **3**, 49–67.

FINNEY, D. J. (1947). The significance of associations in a square point lattice. *J.R. Statist. Soc. Suppl.* **9**, 99–103.

KRISHNA IYER, P. V. (1949). The first and second moments of some probability distributions arising from points on a lattice and their application. *Biometrika*, **36**, 135–41.

KRISHNA IYER, P. V. (1950). The theory of probability distributions of points on a lattice. *Ann. Math. Statist.* **21**, 198–217.

LEGG, J. T. (1953). The natural spread of nettlehead, a virus disease of hops. *Rep. E. Malling Res. Sta. for 1952*, pp. 123–7.

MORAN, P. A. P. (1948). The interpretation of statistical maps. *J.R. Statist. Soc.* B, **10**, 243–51.

PEARSON, E. S. (1947). The choice of statistical tests illustrated on the interpretation of data classed in a 2 × 2 table. *Biometrika*, **34**, 139–67.

SUKHATME, B. V. (1949). Random association of points on a lattice. *J. Indian Soc. Agric. Statist.* **2**, 60–85.

TODD, H. (1940). Note on random associations in a square point lattice. *J.R. Statist. Soc. Suppl.* **7**, 78–82.

VAN DER PLANK, J. E. (1946). A method for estimating the number of random groups of adjacent diseased plants in a homogeneous field. *Trans. Roy. Soc. S. Afr.* **31**, 269.

TESTS OF SIGNIFICANCE FOR CONCURRENT REGRESSION LINES

By E. J. WILLIAMS

Commonwealth Scientific and Industrial Research Organization, Melbourne

The fitting of a straight regression line to a set of observations on two variates is a straightforward application of the method of least squares. Where there are several sets of such data, it is of interest to fit regression lines for each set, and to determine whether or not the slopes of the lines differ more than would be expected due to chance variations. Where the slopes of a set of lines do not differ significantly, the interpretation of the data is particularly straightforward, since the comparison of the sets may be made simply by comparing the constant terms of the regression equations. In fact, if these terms also do not differ significantly when a single slope is used for all lines, then the regression may be regarded as completely homogeneous, and a single equation used to represent the relationship for all the data.

The case of parallel regression lines is particularly important in many applications; for example, the comparison of two materials by means of biological assay can be carried out satisfactorily only when the regression of response (suitably measured) on dosage is linear, and has the same slope for each material. The relative potency is then measured as the distance between the parallel lines, measured in a direction parallel to the dosage axis. When the regression lines are not parallel the interpretation is difficult or impossible.

Another case which appears to be important but has not received so much attention is the case in which the regression lines, instead of being parallel, are concurrent. This type of effect is likely to occur when different materials produce different rates of response, while all give the same response at some fixed level. This level in practice is often not zero, and often not exactly known. When three or more regression lines are concurrent, it is possible to measure the difference between the effects of different materials or treatments by means of the ratios of the distances between the lines measured along an ordinate. It will be seen that parallel regression lines are simply a special case of concurrent lines; in fact, the ratio of the effectiveness of different materials may be measured in precisely the same way as for concurrent lines, that is, by the distances between the intercepts on a transversal.

Tocher (1952) has recently described situations in which it is of interest to know whether the regression lines for different sets of experimental results are concurrent. He has given a test of significance of the hypothesis of concurrence, and a method of determining fiducial limits for the abscissa (i.e. the value of the independent variable) of the point of concurrence. The purpose of this note is to give what is considered to be a more satisfactory solution of these two problems, as well as to discuss further applications of the fitting of concurrent lines.

When the point of concurrence of the lines is known in advance, or only the ordinate requires estimation, the problem is, as pointed out by Tocher, amenable to standard methods. When the abscissa requires estimation, the sums of squares required for the significance tests given here are found to be latent roots of a certain matrix. The analysis and tests of significance are found to parallel those given (Williams, 1952*b*) for the interpretation of interactions in factorial experiments.

FORMULATION OF THE PROBLEM

Our notation differs from that of Tocher, which is not always convenient for the present discussion. We consider m sets of values of x and y; the pairs of values, n_i in number, in the ith set are denoted by x_i and y_i. We confine ourselves to the special case

$$\left.\begin{aligned}\Sigma x_i &= \bar{x} n_i,\\ \Sigma x_i^2 &= X^2 n_i,\end{aligned}\right\} \tag{1}$$

($\bar{x}$ and X^2 being constants; $X \neq 0$ since the x_i cannot all vanish), which includes the commonly occurring case of all sets of data having the same values of x. Without loss of generality, in the mathematical discussion we take

$$\bar{x} = 0, \tag{2}$$

$$\sum_i \Sigma y_i = 0. \tag{3}$$

We also put

$$\Sigma y_i = n_i \bar{y}_i, \tag{4}$$

$$\Sigma y_i x_i = n_i q_i, \tag{5}$$

$$\sum_i n_i = N, \tag{6}$$

$$\sum_i n_i q_i = Nq, \tag{7}$$

and take the point of concurrence as (ξ, η).

(i) *Ordinate of concurrence given*

We assume that η is known, and only ξ is to be estimated. The regression coefficient for the ith line through (ξ, η) is

$$\begin{aligned} b_i &= \frac{\Sigma(y_i - \eta)(x_i - \xi)}{\Sigma(x_i - \xi)^2} \\ &= \frac{q_i - \xi(\bar{y}_i - \eta)}{X^2 + \xi^2}. \end{aligned} \tag{8}$$

The sum of squares for regressions through (ξ, η), with m degrees of freedom, is accordingly

$$\begin{aligned} S &= \sum_i n_i b_i^2 (X^2 + \xi^2) \\ &= \frac{\sum_i n_i q_i^2 - 2\xi \sum_i n_i q_i (\bar{y}_i - \eta) + \xi^2 \sum_i n_i (\bar{y}_i - \eta)^2}{X^2 + \xi^2}. \end{aligned} \tag{9}$$

Since the total sum of squares about η is independent of ξ, to estimate ξ we may equally well maximize S as minimize the residual sum of squares. Put

$$\frac{\Sigma n_i q_i^2}{X^2} = A, \quad \frac{\Sigma n_i q_i (\bar{y}_i - \eta)}{X} = B, \quad \Sigma n_i (\bar{y}_i - \eta)^2 = C.$$

Then (9) gives

$$(X^2 + \xi^2) S = AX^2 - 2BX\xi + C\xi^2, \tag{9'}$$

and S is maximized if

$$\xi S = -BX + C\xi \tag{10}$$

or

$$XS = AX - B\xi. \tag{10'}$$

From (10) and (10′) we see that the maximized S must be the larger latent root of the matrix

$$\begin{bmatrix} A & -B \\ -B & C \end{bmatrix}. \tag{11}$$

The two latent roots, S_1 and S_2, may be given a practical interpretation. S_1 is the sum of squares for regressions and the point of concurrence, and thus has $m+1$ degrees of freedom. S_2, with $m-1$ degrees of freedom, is the sum of squares representing departures from concurrence. Corresponding to the latent root S_1, we have, from (10),

$$\xi = -\frac{\sum_i n_i q_i(\bar{y}_i - \eta)}{S_1 - \sum_i n_i(\bar{y}_i - \eta)^2}$$

$$= -\frac{XB}{S_1 - C}. \tag{12}$$

From the form of the matrix we see that S_2 is zero if and only if

$$q_i \propto \bar{y}_i - \eta, \tag{13}$$

and from (12), that in such cases the coefficient of proportionality is $-X^2/\xi$. Then $b_i = q_i/X^2$, from (8), and the unrestricted regression lines for each set all pass through the point of concurrence. In particular, there is no departure from concurrence when either all the q_i are zero, so that the lines are parallel, or all the $\bar{y}_i = \eta$, so that the point of means is the point of concurrence.

(ii) *Ordinate of concurrence to be estimated*

When η is not given, it may be so chosen as to minimize the residual sum of squares. We then find

$$\eta = \xi q/X^2. \tag{14}$$

But q/X^2 is the average unrestricted regression coefficient for all the data; therefore, as might be expected, the point of concurrence lies on the line of mean regression when η is estimated from the data. It is therefore appropriate to consider departures from mean regression only in testing the significance of departures from concurrence.

The sum of squares for departures of individual regressions from mean regression is

$$S' = \left(\sum_i n_i b_i^2 - Nb^2\right)(X^2 + \xi^2)$$

$$= \frac{\sum_i n_i(q_i - q)^2 - 2\xi \sum_i n_i \bar{y}_i(q_i - q) + \xi^2 \sum_i n_i \bar{y}_i^2}{X^2 + \xi^2}$$

$$= \frac{A'X^2 - 2B'X\xi + C'\xi^2}{X^2 + \xi^2} \quad \text{(say)}, \tag{15}$$

with $m-1$ degrees of freedom. As with S, the maximum value of S' is found as the larger latent root, with m degrees of freedom, of

$$\begin{bmatrix} A' & -B' \\ -B' & C' \end{bmatrix}. \tag{16}$$

Again, the smaller latent root, with $m-2$ degrees of freedom, provides a test of departures from concurrence. The effect of having to estimate η is merely to replace the individual means and regressions by departures from overall means and regressions, and to reduce by one the degrees of freedom.

TESTS OF SIGNIFICANCE

In testing the significance of the differences among regression lines for several sets of data, the usual approach is to examine, first, whether the slopes of the lines differ significantly, and secondly, whether, if the slopes are assumed not to differ, the distances between the parallel lines fitted to the data differ significantly. For the present purpose, the two aspects are viewed in a different way. It is seen from the above discussion that the sum of the two latent roots involves both the differences among regressions and the differences among means, but that the larger latent root, representing differences among regression lines, may involve one or other aspect alone.

The testing of significance of the latent roots of a matrix has been discussed elsewhere (see, for example, Williams, 1952*a*, *b*, and references). The kind of test which seems appropriate to the present problem is one which, first, tests the validity of the assumed model, i.e. whether or not the departure from concurrence is significant, and secondly, gives fiducial limits for ξ. We first consider case (i) above, with η known. Corresponding to a given value of ξ we may determine S, the sum of squares for regressions through (ξ, η), then calculate the quantities

$$v_1 = \frac{(S_1 - S)(S - S_2)}{S} \tag{17}$$

and

$$v_2 = \frac{S_1 S_2}{S}. \tag{18}$$

These are distributed as sums of squares with 1 and $m-1$ degrees of freedom respectively. v_2 gives an exact test of departures from concurrence, corresponding to ξ, and v_1 a test of concordance of the point of concurrence (ξ, η) with the data. Provided v_2 is not significant, v_1 may be used to give a fiducial limit for S at any given probability level, from (17).

It should be emphasized that, although it is often still possible formally to determine fiducial limits for ξ even when v_2 is significant, such fiducial limits for the abscissa of concurrence are irrelevant, since the hypothesis of concurrence is no longer tenable. It is therefore important to test v_2 first before considering the estimation of ξ.

The corresponding fiducial limits for ξ are derived from equations (9) and (17), and are most conveniently given as follows:

$$\xi^2[S - \sum_i n_i(\bar{y}_i - \eta)^2] + 2\xi \sum_i n_i q_i(\bar{y}_i - \eta) + X^2 S - \sum_i n_i q_i^2 = 0,$$

$$\xi = -\frac{\sum_i n_i q_i(\bar{y}_i - \eta) \pm X\sqrt{(Sv_1)}}{S - \sum_i n_i(\bar{y}_i - \eta)^2}$$

$$= -\frac{X(B \pm \sqrt{(Sv_1)})}{S - C}. \tag{19}$$

Alternatively, since v_1 is a sum of squares with 1 degree of freedom, its numerator may be expressed as the square of a rational function of ξ. Since, from (12), v_1 vanishes when

$$\xi = -\frac{XB}{S_1 - C}$$

or

$$\xi = \frac{XB}{C - S_2},$$

it appears that

$$Sv_1 = \frac{[(S_1-C)\,\xi+XB]^2\,[(C-S_2)\,\xi-XB]^2}{B(X^2+\xi^2)^2}$$

$$= \frac{[\xi^2B+\xi X(C-A)-X^2B]^2}{(X^2+\xi^2)^2}. \tag{20}$$

The non-linear analysis of variance may be set out as follows:

	Degrees of freedom	Sums of squares
Point of concurrence	1	$\dfrac{(S_1-S)(S-S_2)}{S}$
Regressions through point of concurrence	m	S
Departures from concurrence	$m-1$	$\dfrac{S_1S_2}{S}$
Total between lines	$2m$	S_1+S_2
Residual	$N-2m$	by difference
Total	N	$\sum_i \sum (y_i-\eta)^2$

When $m=1$, $S_2=0$, so that the test for a given point of concurrence reduces to a test of significance of the difference S_1-S.

For case (ii), when η is to be estimated, a similar approach may be used. Here, however, we require to determine *simultaneous* fiducial limits for ξ and η. The limits will be given as a closed curve surrounding the optimum point.

If the variation of η about the value given by (14) is not considered, we may calculate the statistics

$$v_1' = \frac{(S_1'-S')(S'-S_2')}{S'} \tag{21}$$

and

$$v_2' = \frac{S_1'S_2'}{S'}, \tag{22}$$

with 1 and $m-2$ degrees of freedom respectively. As for the case when η is known, these two statistics may be used to test departures from the abscissa of concurrence ξ, and departures from concurrence, respectively.

Now for the departure of a given value of η from its optimum value $\xi q/X^2$, the sum of squares, with 1 degree of freedom, is

$$\frac{N(\eta X^2-\xi q)^2}{X^2(X^2+\xi^2)}. \tag{23}$$

For a simultaneous test of significance of the concordance of ξ and η with the data we may therefore take

$$v_1'+\frac{N(\eta X^2-\xi q)^2}{X^2(X^2+\xi^2)}$$

$$= S_1'+S_2'-S'-\frac{S_1'S_2'}{S'}+\frac{N(\eta X^2-\xi q)^2}{X^2(X^2+\xi^2)}$$

$$= S_1+S_2-S-\frac{S_1'S_2'}{S'}$$

$$= v_1+v_2-v_2'. \tag{24}$$

This is a sum of squares with 2 degrees of freedom. The set of values of ξ and η giving a constant value to (24) are the required fiducial limits for the position of the point of concurrence.

COMPARISON WITH TOCHER'S RESULTS

The present paper gives significance tests which are exact, and more detailed than those given by Tocher. In fairness it should be remarked, however, that we have treated only the special case defined by (1), and that there are unlikely to be exact tests for the general case treated by Tocher. As there are similarities between the two approaches, it is of interest to compare them. For convenience, we rewrite some of Tocher's results in our notation; his equation numbers will be distinguished by the suffix T.

Tocher first tests the significance of the restraint that the lines be concurrent at a point with abscissa ξ. The criterion is a sum of squares with m degrees of freedom, which is identical with our

$$S_1 + S_2 - S = v_1 + v_2. \tag{25}$$

Its minimum value, which Tocher takes as critical for the test (14T), is clearly S_2, which is reached when $S = S_1$, and ξ has its optimum value. S_2, however, is not distributed as a sum of squares with m degrees of freedom, as Tocher assumes; nor would a reduction of degrees of freedom to $m-1$ be correct, since the equation for optimum ξ is not linear. Tocher's test is, in fact, conservative, as he intends; that is, the significance of departures from concurrence tends to be underestimated (see Williams, 1952*a*, p. 23).

The optimum value of ξ is that for which $S = S_1$. For fiducial limits, Tocher uses the values of ξ corresponding to a value of S making $v_1 + v_2$ significant at a given probability level (15T). The values of ξ are in fact given by (19), with the appropriate value of S. We propose that the setting of limits be done in two stages. First, corresponding to any ξ, v_2, representing departures from concurrence, is tested for significance. If v_2 is not significant, fiducial limits for ξ are determined from the value of v_1 significant at a given probability level. If v_2 is significant, the hypothesis of concurrence must be rejected, and no meaning can be attached to a point of concurrence.

Either method of setting fiducial limits is theoretically correct. However, it is considered that there is a conceptual advantage in distinguishing between the test of the hypothesis of concurrence and the test of a given point of concurrence. When v_2 is small, the use of v_1 also gives narrower fiducial limits than $v_1 + v_2$. This point is illustrated in the numerical example given below.

NUMERICAL EXAMPLES

(i) *Re-examination of Tocher's data*

In order to compare the results given above with Tocher's, we now carry out a comparative analysis of his results. It is noted that he takes $\eta = 0$. From his data (p. 115) we find

$$\begin{aligned}
A &= 0{\cdot}1013\,5806\,6,\\
B &= 0{\cdot}9224\,9221\,3,\\
C &= 8{\cdot}3960\,7140\,0,\\
S_1 + S_2 &= 8{\cdot}4974\,2946\,6,\\
S_1 S_2 &= 0{\cdot}0000\,1767\,7,\\
S_1 &= 8{\cdot}4974\,2738\,6,\\
S_2 &= 0{\cdot}0000\,0208\,0.
\end{aligned}$$

Residual mean square $= 0{\cdot}9223\,1 \times 10^{-6}$ (6 d.f.).

Since the 5 and 1 % points of F, with 1 and 6 degrees of freedom, are 5·9874 and 13·745 respectively, a sum of squares with 1 degree of freedom, to be significant, must exceed

$$5{\cdot}5222 \times 10^{-6} \quad (5\ \%\ \text{level}),$$

or

$$12{\cdot}677 \times 10^{-6} \quad (1\ \%\ \text{level}).$$

Since, for all values of S considered henceforth, the value of v_2 is not significant, departures from concurrence will be assumed non-existent.

We now determine fiducial limits for S, and hence for ξ, from the equation

$$v_1 = 8{\cdot}4974\,2946\,6 - S - \frac{0{\cdot}0000\,1767\,7}{S} = 5{\cdot}5222 \times 10^{-6}$$

and

$$= 12{\cdot}677 \times 10^{-6}.$$

The fiducial limits are as follows:

	Present method		Tocher's method	
	S	ξ	S	ξ
	8·4974		8·4974	
Optimum value	27386	981·5	27386	981·5
5 %	21864	970 and 993	17899	968 and 995
1 %	14709	964 and 999	07234	961 and 1002

The limits by Tocher's method, given for comparison, are found from (15T), or from equation (19) with the appropriate value of S. These ranges are 17 % wider than those from the present method, owing to the inclusion of v_2 in the determination of S.

(ii) *Data requiring estimated ordinate*

In a study of the effect of basis weight on the properties of laboratory-made sheets of paper, three batches of pulp were taken and beaten for different lengths of time. From each batch, sheets of six different basis weights were then made, and mechanical tests carried out on them. The burst strength results from this experiment are set out in Table 1, together with values calculated from them.

Table 1. *Burst strength results*

Basis weight	x	Beating (revs of Lampen Mill)			Total
		1125 $i = 1$	4500 $i = 2$	12730 $i = 3$	
10	−5	1·73	1·98	3·11	
20	−3	4·99	9·69	12·83	
30	−1	8·74	17·26	22·69	
40	1	12·60	24·52	32·45	
50	3	17·04	31·36	48·22	
60	5	21·18	43·29	59·94	
$6\bar{y}_i$		66·28	128·10	179·24	373·62
$6q_i$		137·26	278·82	400·08	816·16

Table 2. *Test of departures of regression lines from origin*

	Degrees of freedom	Sums of squares	Mean square
Correction for mean	3	8821·604	
Regression through means	3	3666·356	
Sum	6	12487·960	
Regressions through origin	3	12311·932	
Departures from origin	3	176·028	58·676**
Residual	12	28·554	2·3795

** Significant at 1 % level.

A preliminary analysis, as given in Table 2, shows that the three regression lines of burst strength on basis weight depart significantly from the origin. In order to compare the trends for different amounts of beating it would be desirable to base the comparison on regression lines through some other common point, if such a model were concordant with the data. The common point would represent the basis weight for which the value of burst factor was independent of beating. The ordinate as well as the abscissa of this point will be estimated.

For the independent variable we take

$$x = \frac{\text{basis weight} - 35}{5},$$

so that x has the values $-5, -3, -1, 1, 3, 5$, and $X^2 = \frac{35}{3}$. The quantities required for the analysis are given in Table 1.

The mean burst strength is 20·76, and the mean regression coefficient on x is

$$b = \frac{816{\cdot}16}{210} = 3{\cdot}886,$$

so that, by (14), the optimum value of the ordinate corresponding to ξ is

$$\eta = 3{\cdot}886\xi + 20{\cdot}76.$$

The matrix (16) is

$$\begin{bmatrix} 494{\cdot}36940 & -726{\cdot}07978 \\ -726{\cdot}07978 & 1066{\cdot}49853 \end{bmatrix},$$

whence

$$S_1' = 1560{\cdot}83437,$$
$$S_2' = 0{\cdot}03356.$$

The optimum value of (ξ, η) is $(-5{\cdot}02, 1{\cdot}26)$ from (12) and (14). Departures from concurrence are clearly not significant, so attention may be confined to $v_1 + v_2 - v_2'$ and the determination of the fiducial boundary for (ξ, η). From the residual mean square given in Table 2, the values for $v_1 + v_2 - v_2'$ required for significance may be determined as

$$18{\cdot}490 \quad (5\,\%\ \text{level}),$$

and

$$32{\cdot}964 \quad (1\,\%\ \text{level}).$$

The equation for the fiducial boundary is, then,

$$\frac{(726{\cdot}1\xi^2+1954\xi-8471)^2}{(\xi^2+\frac{35}{3})(1066\xi^2-4960\xi+5768)}+\frac{210(\eta-3{\cdot}886\xi-20{\cdot}76)^2}{(\xi^2+\frac{35}{3})}=18{\cdot}490 \quad \text{and} \quad 32{\cdot}964.$$

Of particular interest are the maximum ranges for ξ and η. The maximum range for ξ is given above when η does not depart from the value given by the mean regression line. Then, from (19),

$$S=1542{\cdot}34, \quad \xi=-6{\cdot}4 \text{ and } -4{\cdot}0 \quad (5\,\% \text{ level}),$$

$$S=1527{\cdot}87, \quad \xi=-7{\cdot}0 \text{ and } -3{\cdot}7 \quad (1\,\% \text{ level}).$$

The maximum range for η is given when ξ has its optimum value $-5{\cdot}02$; negative values being inadmissible, the limits are found to be

$$0 \quad \text{and} \quad 3{\cdot}1 \quad (5\,\% \text{ level}),$$

$$0 \quad \text{and} \quad 3{\cdot}7 \quad (1\,\% \text{ level}).$$

CONCLUSION

The main emphasis of this paper has been to show that the concepts developed for tests of significance in multivariate analysis and for interactions may be carried over to dealing with a problem of concurrent regressions. It would clearly be possible to extend this work to regression analysis with two or more independent variates. The quantities required for testing significance would then, for k variates, come out as latent roots of a $(k+1)\times(k+1)$ matrix. Such a generalization, however, appears to lack practical interest.

REFERENCES

TOCHER, K. D. (1952). On the concurrence of a set of regression lines. *Biometrika*, **39**, 109.

WILLIAMS, E. J. (1952*a*). Some exact tests in multivariate analysis. *Biometrika*, **39**, 17.

WILLIAMS, E. J. (1952*b*). The interpretation of interactions in factorial experiments. *Biometrika*, **39**, 65.

APPROXIMATE CONFIDENCE INTERVALS

II. MORE THAN ONE UNKNOWN PARAMETER

By M. S. BARTLETT

University of Manchester

1. Retrospect

In a previous paper (Bartlett, 1953; this will be referred to as I) it was shown that the approximate confidence interval obtained from the asymptotic normal distribution for $\partial L/\partial\theta$, where $L \equiv \log p$ is the log likelihood depending on a single unknown parameter θ, could usually be improved by the use of skewness or further moment corrections. The purpose of the present paper is to examine the further problems that arise in the case of more than one unknown parameter. It is perhaps advisable to point out that while the direct use of quantities like L and its derivatives permits a fairly flexible sampling basis, the asymptotic normal distribution remains the basis of the approximations. We may note (see § 7) the application of the proposed method to non-classical problems such as time-series analysis. It could not, however, be used in problems for which the central limit theorem did not apply. This remark appears relevant if we consider (see §§ 5 and 6) some of the more controversial problems involving nuisance parameters, as R. A. Fisher's fiducial probability solutions have been given by some writers (see, for example, Barnard, 1950) an approximate frequency interpretation by means of special (and in my opinion rather artificial*) sampling interpretations.

2. Simultaneous moments

The systematic method of obtaining relations between the moments of $\partial L/\partial\theta$, $\partial^2 L/\partial\theta^2$, etc., is obviously extensible to the case of several unknowns θ_i $(i = 1, 2, \ldots)$. The derivation of the more general formulae rapidly becomes rather tedious, and I shall only record the formulae up to the third-order moments of $\partial L/\partial\theta_i$. On the usual assumptions about differentiation with respect to θ_i commuting with the integral sign, we may write (cf. I)

$$1 = \int p(\theta_i + \tau_i) = \int p(\theta_i) \exp\left\{\Sigma_i \tau_i \frac{\partial L}{\partial\theta_i} + \tfrac{1}{2}\Sigma_i \Sigma_j \tau_i \tau_j \frac{\partial^2 L}{\partial\theta_i \partial\theta_j} \cdots\right\},$$

whence

$$E\left\{\frac{\partial L}{\partial\theta_i}\right\} = 0, \quad E\left\{\frac{\partial L}{\partial\theta_i}\frac{\partial L}{\partial\theta_j}\right\} = E\left\{-\frac{\partial^2 L}{\partial\theta_i\partial\theta_j}\right\},$$

$$E\left\{\frac{\partial L}{\partial\theta_i}\frac{\partial L}{\partial\theta_j}\frac{\partial L}{\partial\theta_k}\right\} = E\left\{-\frac{\partial^3 L}{\partial\theta_i\partial\theta_j\partial\theta_k}\right\} - E\left\{\frac{\partial L}{\partial\theta_i}\frac{\partial^2 L}{\partial\theta_j\partial\theta_k}\right\} - E\left\{\frac{\partial L}{\partial\theta_j}\frac{\partial^2 L}{\partial\theta_i\partial\theta_k}\right\} - E\left\{\frac{\partial L}{\partial\theta_k}\frac{\partial^2 L}{\partial\theta_i\partial\theta_j}\right\}.$$

Differentiating the second-order equation, we see further that

$$\frac{\partial}{\partial\theta_i} E\left\{\frac{\partial^2 L}{\partial\theta_j\partial\theta_k}\right\} + E\left\{\frac{\partial L}{\partial\theta_i}\frac{\partial L}{\partial\theta_j}\frac{\partial L}{\partial\theta_k}\right\} + E\left\{\frac{\partial^2 L}{\partial\theta_i\partial\theta_j}\frac{\partial L}{\partial\theta_k}\right\} + E\left\{\frac{\partial^2 L}{\partial\theta_i\partial\theta_k}\frac{\partial L}{\partial\theta_j}\right\} = 0.$$

---

* It is not my purpose to cloud the present discussion with any controversial fog. I would, however, note that the sampling procedure suggested by Barnard in the Behrens-Fisher problem, if deliberately carried out, would lead approximately to the solution claimed, but at the cost of throwing away the information on the variances available in the expected further observations. It seems to me that we should rarely want to do this unless the resulting loss of accuracy in our inference on the mean difference was negligible, but this is just the limiting situation for which we do not require small-sample theory at all.

From these last two equations, we find that

$$E\left\{\frac{\partial L}{\partial\theta_i}\frac{\partial L}{\partial\theta_j}\frac{\partial L}{\partial\theta_k}\right\} = 2E\left\{\frac{\partial^3 L}{\partial\theta_i\partial\theta_j\partial\theta_k}\right\} + \frac{\partial}{\partial\theta_i}I_{jk} + \frac{\partial}{\partial\theta_j}I_{ik} + \frac{\partial}{\partial\theta_k}I_{ij}, \tag{1}$$

where $I_{ij} \equiv E\{-\partial^2 L/\partial\theta_i\partial\theta_j\}$. Formulae of the same order involving less than three parameters may be obtained from (1) by contraction (i.e. making two or more θ's identical).

3. CONFIDENCE REGIONS

As already noted in I, the extension from approximate confidence intervals to confidence regions is in principle straightforward. Thus for two parameters θ_1 and θ_2 our first large-sample approximation is to consider the joint quantities $\partial L/\partial\theta_1$, $\partial L/\partial\theta_2$ as normal with zero means and known covariance-matrix. From these we may, for example, construct an approximate χ^2 expression, writing this (for convenience later) in the unsymmetric form

$$\left(\frac{\partial L}{\partial\theta_1}\right)^2 \Big/ I_{11} + \left(\frac{\partial L}{\partial\theta_2} - \frac{I_{12}}{I_{11}}\frac{\partial L}{\partial\theta_1}\right)^2 \Big/ \left(I_{22} - \frac{I_{12}^2}{I_{11}}\right). \tag{2}$$

Values of θ_1 and θ_2 for which the expression in (2) remains less than χ_0^2, the critical value of χ^2 for two degrees of freedom, define an approximate confidence region. An even simpler expression asymptotically equivalent to (2) is $-2[L - L_{\max}]$, discussed more fully later.*

It is, perhaps, hardly necessary to consider any skewness corrections to (2), for in it only *squares* of the approximately normal quantities are involved. Any skewness corrections would merely help to locate the minimum and maximum 'admissible' values of θ_1 and θ_2 more symmetrically (in terms of probability contributions to the total probability of error). However, to illustrate the procedure these corrections will be deduced. Consider

$$\left.\begin{aligned}
T_1(\theta_1,\theta_2) &\equiv \frac{\partial L}{\partial\theta_1} + \lambda_1\left[\left(\frac{\partial L}{\partial\theta_1}\right)^2 - I_{11}\right],\\
T_2(\theta_1,\theta_2) &\equiv \frac{\partial L}{\partial\theta_2} - \frac{I_{12}}{I_{11}}\frac{\partial L}{\partial\theta_1} + \lambda_{20}\left[\left(\frac{\partial L}{\partial\theta_1}\right)^2 - I_{11}\right] + \lambda_{11}\left[\frac{\partial L}{\partial\theta_1}\frac{\partial L}{\partial\theta_2} - I_{12}\right] + \lambda_{02}\left[\left(\frac{\partial L}{\partial\theta_2}\right)^2 - I_{22}\right].
\end{aligned}\right\} \tag{3}$$

In order to replace the random variables in (2) by more nearly normal variables, we require the third-order cumulants of T_1 and T_2 to be zero to the next-order approximation. This gives

$$\left.\begin{aligned}
&\kappa_{30} + 6\lambda_1 I_{11}^2 = 0 \quad (\text{cf. I}),\\
&\left(\kappa_{21} - \frac{I_{12}}{I_{11}}\kappa_{30}\right) + 2\lambda_{20}I_{11}^2 + 2\lambda_{11}I_{11}I_{12} + 2\lambda_{02}I_{12}^2 = 0,\\
&\left(\kappa_{12} - 2\frac{I_{12}}{I_{11}}\kappa_{21} + \frac{I_{12}^2}{I_{11}^2}\kappa_{30}\right) + 2\lambda_{11}(I_{11}I_{22} - I_{12}^2) + 4\lambda_{02}I_{12}\left(I_{22} - \frac{I_{12}^2}{I_{11}}\right) = 0,\\
&\left(\kappa_{03} - 3\frac{I_{12}}{I_{11}}\kappa_{12} + 3\frac{I_{12}^2}{I_{11}^2}\kappa_{21} - \frac{I_{12}^3}{I_{11}^3}\kappa_{30}\right) + 6\lambda_{02}\left(I_{22} - \frac{I_{12}^2}{I_{11}}\right)^2 = 0.
\end{aligned}\right| \tag{4}$$

The last three equations may be solved successively from the bottom to the top equation; in the particular case when $I_{12} = 0$ they become

$$\kappa_{21} + 2\lambda_{20}I_{11}^2 = 0, \quad \kappa_{12} + 2\lambda_{11}I_{11}I_{22} = 0, \quad \kappa_{03} + 6\lambda_{02}I_{22}^2 = 0. \tag{5}$$

---

* It is convenient to defer discussion of this alternative until § 8, as so far no explicit use of maximum-likelihood estimates has been necessary, although we shall find their introduction in general unavoidable from § 4 onwards.

Thus for a normal sample with

$$L = -\frac{n}{2}\log\theta_2 - \frac{n}{2}\frac{(\bar{x}-\theta_1)^2}{\theta_2} - \frac{1}{2}\frac{(n-1)\,s^2}{\theta_2}, \tag{6}$$

so that
$$\frac{\partial L}{\partial\theta_1} = \frac{n(\bar{x}-\theta_1)}{\theta_2}, \quad \frac{\partial L}{\partial\theta_2} = -\frac{n}{2\theta_2} + \frac{(n-1)\,s^2+n(\bar{x}-\theta_1)^2}{2\theta_2^2},$$

$$-\frac{\partial^2 L}{\partial\theta_1^2} = \frac{n}{\theta_2}, \quad -\frac{\partial^2 L}{\partial\theta_1\partial\theta_2} = \frac{n(\bar{x}-\theta_1)}{\theta_2^2}, \quad -\frac{\partial^2 L}{\partial\theta_2^2} = -\frac{n}{2\theta_2^2} + \frac{(n-1)\,s^2+n(\bar{x}-\theta_1)^2}{\theta_2^3},$$

$$\left.\begin{aligned} &I_{11} = \frac{n}{\theta_2}, \quad I_{12} = 0, \quad I_{22} = \frac{n}{2\theta_2^2}, \\ &\kappa_{30} = 0, \quad \kappa_{21} = \frac{n}{\theta_2^2}, \quad \kappa_{12} = 0, \quad \kappa_{03} = \frac{n}{\theta_2^3}, \end{aligned}\right\} \tag{7}$$

we put
$$\lambda_1 = \lambda_{11} = 0, \quad \lambda_{20} = -\frac{1}{2n}, \quad \lambda_{02} = -\frac{2\theta_2}{3n},$$

$$T_1 = \frac{n(\bar{x}-\theta_1)}{\theta_2},$$

$$T_2 = -\frac{n}{2\theta_2} + \frac{(n-1)\,s^2+n(\bar{x}-\theta_1)^2}{2\theta_2^2} - \frac{1}{2n}\left[\frac{n^2(\bar{x}-\theta_1)^2}{\theta_2^2} - \frac{n}{\theta_2}\right] - \frac{2\theta_2}{3n}\left[\left(\frac{\partial L}{\partial\theta_2}\right)^2 - \frac{n}{2\theta_2^2}\right].$$

Now as the crude approximate region is given by

$$\frac{n(\bar{x}-\theta_1)^2}{\theta_2} + \frac{2\theta_2^2}{n}\left(\frac{\partial L}{\partial\theta_2}\right)^2 = \chi_0^2,$$

we may write in the next approximation

$$\begin{aligned} T_2 &= -\frac{n}{2\theta_2} + \frac{(n-1)\,s^2}{2\theta_2^2} + \frac{1}{2\theta_2} + \frac{1}{3\theta_2} - \frac{1}{3\theta_2}\left[\chi_0^2 - \frac{n(\bar{x}-\theta_1)^2}{\theta_2}\right] \\ &= \frac{n-1}{2\theta_2^2}\,[s^2-\theta_2] + \frac{1}{3\theta_2}\left[1 + \frac{n(\bar{x}-\theta_1)^2}{\theta_2} - \chi_0^2\right], \end{aligned} \tag{8}$$

obtaining as the next-order approximation for the confidence region

$$\frac{n(\bar{x}-\theta_1)^2}{\theta_2} + \frac{2\theta_2^2}{n}T_2^2 = \chi_0^2, \tag{9}$$

where T_2 is given by (8). As an obvious check on this result, we may note that if we had initially made use of the facts that $\bar{x}$ and s^2 are sufficient independent statistics for θ_1 and θ_2, we should have arrived at practically the same equation (9) by making use of the normal approximation for s^2 developed in I (as noted in I, this is not the *best* normal approximation for s^2, but the one arrived at by the present systematic procedure). The only difference would be that the coefficient of T_2^2 in (9) would be $2\theta_2^2/(n-1)$, this difference arising from the neglect, justifiable to the order of approximation considered, of the alteration in the variance of T_2 due to the third-order moment corrections.

4. THE CASE OF 'NUISANCE PARAMETERS'

We come now to the more formidable problem of 'nuisance parameters'. To fix our problem, consider first the case of two parameters θ_1 and θ_2, with θ_2 the nuisance parameter. The quantity involving $\partial L/\partial\theta_1$, $\partial L/\partial\theta_2$, which provides an alternative to the maximum-likelihood estimate $\hat{\theta}_1$, is

$$\frac{\partial L}{\partial\theta_1}-\frac{I_{12}}{I_{22}}\frac{\partial L}{\partial\theta_2}, \tag{10}$$

with variance $I_{11}-I_{12}^2/I_{22}\equiv I_{11\cdot 2}$, say. But only in large samples, when the expression in (10) and $\partial L/\partial\theta_2$ become uncorrelated normal variables, with $\partial L/\partial\theta_2$ an effectively sufficient statistic for θ_2, can we expect the standardized variable

$$T\equiv\left(\frac{\partial L}{\partial\theta_1}-\frac{I_{12}}{I_{22}}\frac{\partial L}{\partial\theta_2}\right)\Big/\sqrt{I_{11\cdot 2}} \tag{11}$$

to be completely insensitive to the estimate we substitute for θ_2. In smaller samples this substitution will affect the distribution of T. In principle this effect can be successively allowed for, this being the proposed procedure; the effect of the correction at one stage may be reconsidered, a further correction of a higher order introduced, and so on. As a general method, this is, however, evidently going to become rather intractable after the first stage, for the previous advantages of working directly with the derivatives $\partial L/\partial\theta_1$, $\partial L/\partial\theta_2$ are in general lost.

It is not possible to ignore the sufficiency properties, if any, that exist for θ_2. For if $\partial L/\partial\theta_2$ is equivalent, even in small samples, to a single sufficient statistic for θ_2 (when θ_1 is assumed known), we *can** proceed from T by building up a polynomial series solution in $\partial L/\partial\theta_1$ and $\partial L/\partial\theta_2$ which is *statistically* independent of $\partial L/\partial\theta_2$ and hence *mathematically* independent of θ_2. But if $\partial L/\partial\theta_2$ has not got such sufficiency properties, it can only be considered 'sufficient' for θ_2 if regarded as a random *function* in the parameter θ_2, and statistical independence merely of the random *variable* $\partial L/\partial\theta_2$ would appear to be no longer adequate. I thought at one stage that this difficulty might be overcome by eliminating also the dependence on higher derivatives such as $\partial^2 L/\partial\theta_2^2$, but this may be shown (by a counter-example, see § 6) to be impossible in general. In the general case it seems therefore necessary to fall back on the very cumbersome substitution procedure referred to in the first paragraph above.

Fortunately this substitution, if the maximum-likelihood estimate of θ_2 for given θ_1 is used, only affects the distribution of the statistic T to the order of approximation $O(1/n)$, where n indicates the sample size, a stage higher than the skewness correction, which is in general of $O(1/\sqrt{n})$. For we may write the maximum-likelihood estimate of θ_2 (for given θ_1) approximately as

$$\theta_2+\frac{\partial L}{\partial\theta_2}\Big/I_{22},$$

so that

$$T_e\sim T+\frac{\partial L}{\partial\theta_2}\frac{\partial T}{\partial\theta_2}\Big/I_{22}. \tag{12}$$

The *estimated* value of the second term in (12) is of course zero. Its sampling effect on T_e may, however, be investigated. The first correcting term in $E\{(T_e)^r\}$ is

$$\begin{aligned}rE\left\{T^{r-1}\frac{\partial L}{\partial\theta_2}\frac{\partial T}{\partial\theta_2}\Big/I_{22}\right\}&=E\left\{\frac{\partial L}{\partial\theta_2}\frac{\partial T^r}{\partial\theta_2}\Big/I_{22}\right\}\\&=\frac{\partial}{\partial\theta_2}E\left\{T^r\frac{\partial L}{\partial\theta_2}\Big/I_{22}\right\}-E\left\{T^r\left(\frac{\partial L}{\partial\theta_2}\right)^2\Big/I_{22}\right\}-E\left\{T^r\frac{\partial^2 L}{\partial\theta_2^2}\Big/I_{22}\right\}+E\left\{T^r\frac{\partial L}{\partial\theta_2}\frac{\partial I_{22}}{\partial\theta_2}\Big/I_{22}^2\right\}.\end{aligned}$$

---

* (Added in proof.) I do not mean to imply that we *should*, for in this case the probability of the sample, given $\partial L/\partial\theta_2$, provides a distribution independent of θ_2 (cf. Bartlett, 1936), and the series expansion, illustrated by § 5 on the t-test, is unnecessary except perhaps for finding approximations.

From the asymptotic independence of T and $\partial L/\partial\theta_2$, it is easy to see that all these terms vanish to the first order (for the second and third, we replace $-\partial^2 L/\partial\theta_2^2$ in the third term by I_{22} to the first order, and couple with the second term). Other asymptotically efficient estimates will be equivalent to the same order and may be used if more convenient; their effect at higher stages would have to be investigated separately.

For reference if required, the next term in the expansion of $\delta\theta_2$ gives

$$\delta\theta_2 \sim \frac{\dfrac{\partial L}{\partial\theta_2}}{I_{22}} + \left\{\frac{\dfrac{\partial L}{\partial\theta_2}\left[\dfrac{\partial^2 L}{\partial\theta_2^2}+I_{22}\right]}{I_{22}^2} + \frac{\left(\dfrac{\partial L}{\partial\theta_2}\right)^2 \dfrac{\partial^3 L}{\partial\theta_2^3}}{2I_{22}^3}\right\}, \tag{13}$$

whence

$$T_e \sim T + \frac{\partial L}{\partial\theta_2}\frac{\partial T}{\partial\theta_2}\bigg/ I_{22} + f\frac{\partial T}{\partial\theta_2} + \frac{1}{2}\left(\frac{\partial L}{\partial\theta_2}\right)^2 \frac{\partial^2 T}{\partial\theta_2^2}\bigg/ I_{22}^2, \tag{14}$$

where f is the expression in curly brackets in (13).

A result of some interest which may be derived from an expansion like (13) is

$$E\{\hat{\theta}\} = \theta - \left\{\tfrac{1}{2}E\left\{\frac{\partial^3 L}{\partial\theta^3}\right\} + \frac{\partial I}{\partial\theta}\right\}\bigg/ I^2 + \ldots, \tag{15}$$

where the suffix 2 has been dropped for convenience, as (13) and (15) refer effectively to a single unknown parameter.

While it would be too laborious to carry through the complete details of the general correction to the $O(1/n)$ stage, it may be advisable to indicate how this would proceed, so that the iterative nature of the corrections is made clear. To $O(1/\sqrt{n})$ we have

$$S_e = [T - \tfrac{1}{6}\gamma_1(T^2-1)]_e,$$

where γ_1 is the skewness of T. Owing *both* to the properties of S and to the substitution of the estimate S_e for S, so that S_e is related to S by an expansion similar to that for T_e in (14), we have $O(1/n)$ additions to the mean, variance, third and fourth cumulants of, say, c_1, c_2, c_3 and c_4 respectively. Then a new quantity R correct to $O(1/n)$ is (cf. Cornish & Fisher, 1937)

$$R = S_e - c_1 - \tfrac{1}{2}c_2 S_e - \tfrac{1}{6}c_3(S_e^2-1) - \tfrac{1}{24}c_4(S_e^3 - 3S_e); \tag{16}$$

the effect of further substitution for θ_2 in c_1, c_2, c_3 and c_4 will not affect the distribution of R_e before at least $O(1/n^{\frac{3}{2}})$.

As noted in I, a statement of general conditions could hardly be expected under which the required monotonic relation existed between our final statistical function and θ_1 if a valid confidence interval is to be possible, but except for very small samples we should not expect these higher-order corrections to disturb any such relation holding in the relevant range; this expectation has been supported by numerical examples.

5. The *t*-test and the Behrens-Fisher problem

It seems of some interest to demonstrate the procedure first in the case of the standard *t*-test, even although not only do we not require it here but from the above remarks it does not illustrate the completely general case.

From the expression (6) we obtain for the quantity T defined in (11)

$$T = \frac{\sqrt{n}(\bar{x}-\theta_1)}{\sqrt{\theta_2}}.$$

From the sufficiency property of $\partial L/\partial\theta_2$ in this example, a successive elimination of the higher cumulants between $\partial L/\partial\theta_2$ and the modified series

$$\sqrt{\frac{n}{\theta_2}}\,T_\mu \equiv \sqrt{\frac{n}{\theta_2}}\,T + \mu_{20}\left[\left(\frac{\partial L}{\partial\theta_1}\right)^2 - I_{11}\right] + \mu_{11}\left[\frac{\partial L}{\partial\theta_1}\frac{\partial L}{\partial\theta_2} - I_{12}\right] + \mu_{02}\left[\left(\frac{\partial L}{\partial\theta_2}\right)^2 - I_{22}\right] + \dots$$

will eliminate the dependence of T_μ on θ_2. We choose the μ's also so that T_μ becomes normal. The resulting equations are similar to (5), being

$$\kappa_{30} + 6\mu_{20}I_{11}^2 = 0, \quad \kappa_{21} + 2\mu_{11}I_{11}I_{22} = 0, \quad \kappa_{12} + 2\mu_{02}I_{22}^2 = 0,$$

or $\mu_{20} = 0$, $\mu_{11} = -\theta_2/n$, $\mu_{02} = 0$, so to the next approximation stage

$$T_\mu \equiv \frac{\sqrt{n}(\bar{x}-\theta_1)}{\sqrt{\theta_2}}\left(1 - \frac{1}{2}\left\{\frac{(n-1)\,s^2 + n(\bar{x}-\theta_1)^2}{n\theta_2} - 1\right\}\right)$$

$$\sim \frac{\sqrt{n}(\bar{x}-\theta_1)}{\left\{\dfrac{(n-1)\,s^2 + n(\bar{x}-\theta_1)^2}{n}\right\}^{\frac{1}{2}}}.$$

This result is not of course yet correct to $O(1/n)$, but there is little point in continuing further with the corrections. If we make use of (12), we reach the same result to this order more simply. Higher approximations would merely develop the Fisher (1941) asymptotic expansion for t.

In the case of the Behrens-Fisher problem of the difference of two independent normal variables with separately unknown variances, we have

$$L = -\tfrac{1}{2}\log(\theta_1+\theta_2) - \frac{1}{2}\frac{(x-\theta_3)^2}{\theta_2+\theta_3} - \frac{n_1}{2}\log\theta_1 - \frac{1}{2}\frac{n_1 v_1}{\theta_1} - \frac{n_2}{2}\log\theta_2 - \frac{1}{2}\frac{n_2 v_2}{\theta_2}, \tag{17}$$

where to correspond with the degrees of freedom n_1, n_2 for the estimated variances v_1, v_2, the unknown variances of x_1, x_2 are taken as θ_1 and θ_2, and θ_3 for the unknown true difference corresponding to $x \equiv x_1 - x_2$.

We easily deduce that

$$T = \frac{x-\theta_3}{\sqrt{(\theta_1+\theta_2)}},$$

and if we substitute the most convenient estimates v_1, v_2 for θ_1, θ_2 (these are not the strict maximum-likelihood estimates for given θ_3, but are asymptotically equivalent as n_1, n_2 increase), we arrive at the usual approximate t-quantity (see, for example, Welch (1947)). To reach a result correct to $O(1/n)$ we should require (i) to correct the kurtosis of T (its skewness is obviously zero), (ii) to make the $O(1/n)$ correction corresponding to the effect of the substitution of v_1, v_2 for θ_1, θ_2. There is little point in carrying this through, for it is clear that it would lead to the next order correction in the asymptotic series solution to this problem proposed by Welch (1947). My impression (with which Dr Welch may not necessarily agree) is that his remarkably ingenious solution to this problem may be logically classified with the further asymptotic expansions based on the normal approximation which are the subject-matter of I and the present paper.

If we alternatively tried to extend the first procedure illustrated for the t-test, to cover two nuisance parameters θ_1 and θ_2, we should eliminate the dependence on both $\partial L/\partial\theta_1$ and $\partial L/\partial\theta_2$. It has been verified to the first stage of the approximation procedure that this leads similarly to a consideration of the approximate t-quantity $(x-\theta_3)/\sqrt{(v_1+v_2)}$, but we would not expect it to extend to higher stages for the reasons given in § 4, as $\partial L/\partial\theta_1$ and $\partial L/\partial\theta_2$ do not correspond to a pair of sufficient statistics for θ_1 and θ_2 (for given θ_3).

6. FISHER'S ANALYSIS OF VARIANCE PROBLEM

A simpler problem for which such difficulties may be illustrated is the second problem considered by Fisher (1935) in his first derivation of solutions in terms of simultaneous fiducial probability. This problem may be stated for our present purpose as follows. We have two χ^2-quantities

$$\frac{nv}{\theta_1+\lambda\theta_2},\quad \frac{n_2 v_2}{\theta_2},$$

with degrees of freedom n and n_2, and require a confidence interval for θ_1. Here

$$\begin{aligned}
L &= -\frac{n}{2}\log(\theta_1+\lambda\theta_2)-\frac{n}{2}\left(\frac{v}{\theta_1+\lambda\theta_2}\right)-\frac{n_2}{2}\log\theta_2-\frac{n_2}{2}\frac{v_2}{\theta}\\
&= L(n)+L(n_2),\text{ say,}\\
\frac{\partial L}{\partial\theta_1} &= \frac{n}{2(\theta_1+\lambda\theta_2)^2}[v-(\theta_1+\lambda\theta_2)],\\
\frac{\partial L}{\partial\theta_2} &= \lambda\frac{\partial L}{\partial\theta_1}+\frac{\partial L(n_2)}{\partial\theta_2},\\
I_{11} &= \frac{n}{2(\theta_1+\lambda\theta_2)^2},\quad I_{12}=\lambda I_{11},\quad I_{22}=\lambda^2 I_{11}+\frac{n_2}{2\theta_2^2},
\end{aligned}$$

whence after reduction we obtain

$$T=(v-\lambda v_2-\theta_1)\sqrt{\frac{nn_2}{2[n_2(\theta_1+\lambda\theta_2)^2+\lambda^2 n\theta_2^2]}}. \tag{18}$$

While this problem appears at first sight rather simple, depending on two joint sufficient statistics v and v_2 for θ_1 and θ_2, the difficulty arises from the absence of a sufficient statistic for θ_2 (for given θ_1). Hence we cannot expect it to suffice to modify T to be independent of $\partial L/\partial\theta_2$. In this example we may notice further that it is impossible to make T independent also of, say, $\partial^2 L/\partial\theta_2^2$, for as T, $\partial L/\partial\theta_2$ and $\partial^2 L/\partial\theta_2^2$ are all linear in v and v_2, there is a linear relation between them. Thus we return to the substitution procedure recommended in §4.

It will also be seen in this example that (as in the Behrens-Fisher problem) the use of the normal approximation depends on *two* sample numbers not being too small, in this case n and n_2. The best estimate of θ_2 for given θ_1 is not v_2, but can be obtained from the equation $\partial L/\partial\theta_2=0$ using v_2 as a first approximation. This leads to the estimate

$$\left\{\frac{n_2}{v_2}+\frac{\lambda n(v-\theta_1)}{(\theta_1+\lambda v_2)^2}\right\}\Big/\left\{\frac{n_2}{v_2^2}+\frac{\lambda^2 n}{(\theta_1+\lambda v_2)^2}\right\}. \tag{19}$$

We may not substitute the simpler estimate v_2 (unless $n_2 \gg n$), for though its accuracy is of the same order as the estimate in (19), it is not asymptotically independent of T and hence the effect of the substitution will be of $O(1/\sqrt{n}, 1/\sqrt{n_2})$ and must also be corrected even at the first stage of the correction.

The skewness of T, unaffected to the first order by the substitution, is

$$\gamma_1=8\left\{\frac{(\theta_1+\lambda\theta_2)^3}{n^2}-\frac{\lambda^3\theta_2^3}{n_2^2}\right\}I_{11\cdot 2}^{\frac{3}{2}}, \tag{20}$$

where $I_{11\cdot 2}$ is the expression under the square root in (18).

Thus corrected to $O(1/\sqrt{n}, 1/\sqrt{n_2})$ we may consider our confidence interval for θ_1 given by either of the two equations (cf. I)

$$T - \tfrac{1}{6}\gamma_1(T^2-1) = \pm\mu, \tag{21a}$$

$$T - \tfrac{1}{6}\gamma_1(\mu^2-1) = \pm\mu, \tag{21b}$$

where in (21*a*) and (21*b*) it is understood that the estimate in (19) is written for θ_2. An equation correct to $O(1/n, 1/n_2)$ would be rather complicated, for in addition to a correction for the substitution for θ_2, a correction for kurtosis would be required; this would make the total correction exceedingly cumbersome, as a bias, variance correction and the asymmetry must all be allowed for at this stage (see § 4).

However, from the check made in I on the accuracy of the confidence interval for a single variance, the use of equation (21*a*) or (21*b*) would be not unreasonable for moderate-sized n and n_2.

Table 1

n	χ^2 quantities	θ_2	v	v_2	Lower and upper $P = 0{\cdot}05$ limits for θ_1
12	11·76, 13·43	0	0·980	0	0·55 (0·56), 2·32 (2·25)
		2	1·960	2·240	0, 3·53
		4	2·940	4·476	0, 4·74
	15·24, 5·46	0	1·270	0	0·72 (0·72), 3·00 (2·92)
		2	2·540	0·910	0·91, 5·59
		4	3·810	1·820	0·86, 8·06
6	1·85, 7·80	0	0·308	0	0·14 (0·15), 1·28 (1·13)
		2	0·617	2·600	0, 1·16
		4	0·926	5·196	0, 1·15
	5·64, 4·75	0	0·940	0	0·44 (0·45), 3·92 (3·46)
		2	1·880	1·583	0, 7·09
		4	2·820	3·167	0, 10·32

To illustrate and examine the numerical use of equation (21*b*), it was applied to a few artificial sampling estimates in the case when there are only two observations per group, so that in the analysis of variance table the between-and-within-groups have an equal number of degrees of freedom, n. The true mean square between groups is $2\theta_1+\theta_2$, so to correspond to the notation above (with $\lambda = \frac{1}{2}$) we take as v one-half of the between-groups mean square and v_2 as the within-groups mean square. Two cases were taken for $n = 12$, and two for $n = 6$.* In addition, the same effective χ^2 quantities were used when comparing the results for two different values of θ_2 (2 and 4), so that the extent to which the inference depended on θ_2 could be more easily seen. On this last point, however, it should be recalled that the statistic used is only independent of our error in estimating θ_2 up to and including $O(1/\sqrt{n})$ in *distribution*, and in any particular case (for which the *error* in θ_2 will alter with θ_2 even for the same χ^2) will still change to an extent of $O(1/\sqrt{n})$.

Moreover, the information on θ_1 is of course less as θ_2 increases, so the width of the interval should in any case tend to increase with θ_2. For comparison the case $\theta_2 = 0$ has been included, as this reduces to the single variance problem discussed in I, and the exact confidence interval in this case can also be given in Table 1 (in brackets). The approximate intervals for $\theta_2 \neq 0$ were obtained from equation (21*b*) graphically, plotting against θ_1 for the lower limit and $1/\theta_1$ for the upper limit (these scales are indicated by the form of the equation at the permissible extreme points $\theta_1 \to 0$ and $1/\theta_1 \to 0$). The true value of θ_1 throughout was unity.

---

* It had originally been intended to examine ten cases each, but as this would still be far from sufficient to test the approximation adequately, it was later thought hardly worth while. The sample quantities v and v_2 made use of were partly selected from the larger set as interesting combinations to try out, and thus are not entirely random.

The lower limit for θ_1 will often be the minimum permissible value zero, rather than the value corresponding to $\mu = -1{\cdot}645$, which may well not be reached, especially for small n and large θ_2. This is not unexpected, as it is well known that even the strict maximum-likelihood estimate $v - \lambda v_2$ in this problem may often turn out to be negative; it implies, however, that we are not making full use of the $0{\cdot}05$ risk of error at the lower value, so that it seems useful to think of the error risks for the lower and upper limits separately.*

7. Application to time-series analysis

It has previously been noted (Bartlett, 1952, at end of paper) that the technique may be applied to unknown parameters in time-series. We are now in a position to consider further the solution proposed there for the auto-correlation of a normal Markoff process, in discrete time, and to give a modified solution corrected to $O(1/\sqrt{n})$. The quantity L is, apart from the end-correction (which will not affect our solution until $O(1/n)$)

$$L = -\frac{n}{2}\log\sigma_Y^2 - \frac{1}{2}\sum_{r=1}^{n} Y_r^2/\sigma_Y^2, \tag{22}$$

where

$$X_r - \beta X_{r-1} = Y_r.$$

(For simplicity we assume, as before, that $E\{Y\} = 0$. In (22) the observation X_0 has been discarded.) We have†

$$\frac{\partial L}{\partial\beta} = \sum_{r=1}^{n}(X_r - \beta X_{r-1})\,X_{r-1}/\sigma_Y^2,$$

$$I_{11} = \frac{n}{1-\beta^2}, \quad I_{12} = 0, \quad I_{22} = \frac{n}{2\sigma_Y^4},$$

where $\theta_1 \equiv \beta$, $\theta_2 \equiv \sigma_Y^2$. We have further

$$\frac{\partial^3 L}{\partial\beta^3} = 0, \quad \frac{\partial I_{11}}{\partial\beta} = \frac{2n\beta}{(1-\beta^2)^2},$$

whence

$$E\left\{\left(\frac{\partial L}{\partial\beta}\right)^3\right\} = \frac{6n\beta}{(1-\beta^2)^2}.$$

We note finally that

$$\hat{\sigma}_Y^2 = \sum_{r=1}^{n} X_r^2(1-\beta^2)/n. \tag{23}$$

Hence our approximate confidence interval for β is determined from the equation

$$\frac{\sum_{r=1}^{n}(X_r - \beta X_{r-1})\,X_{r-1}/\hat{\sigma}_Y^2}{\sqrt{\{n/(1-\beta^2)\}}} - \frac{\beta}{\sqrt{\{(1-\beta^2)\,n\}}}(\mu^2 - 1) = \pm\mu. \tag{24}$$

---

* In the case $n = n_2$ the expression in (21) at $\theta_1 = 0$ is $\sqrt{n}(v - \lambda v_2)/(v + \lambda v_2)$, which must be greater than μ for a non-zero lower limit. When $\lambda v_2/v < \frac{1}{3}$, it may be shown that the monotonic relation with θ_1 breaks down near $\theta_1 = 0$, but of course *if* θ_1 is near zero $\lambda v_2/v \sim 1 + O(1/\sqrt{n})$, so that no difficulty should arise within the 'admissible' range of θ_1 except perhaps for very small n.

† In equation (25) of the paper cited, σ_Y^2 was inadvertently omitted, but this did not affect the deduced confidence interval equation (26).

It is interesting to notice the appearance of a non-zero skewness correction, in contrast with the classical regression problem.

When we are able to assume normal residuals (or some other complete specification of the process) similar methods can clearly be applied in other time-series problems.

8. RELATION WITH GOODNESS OF FIT TESTS

It was pointed out in §3 that an alternative to the approximate χ^2 expression in $\partial L/\partial\theta_i$ would be to use $-2(L-L_{\text{max.}})$, where $L_{\text{max.}}$ denotes the maximum of L for choice of the unknown θ_i. This last expression is more familiar in the use of the χ^2 test as a goodness of fit of hypothetical values of the θ_i, but the two uses are of course closely connected. In this final section we shall demonstrate *to the next order of approximation* ($O(1/\sqrt{n})$) the asymptotic equivalence (in distribution) of this alternative χ^2 approximation,* equivalent to the likelihood ratio λ and given by

$$-2\log\lambda = -2(L-L_{\text{max.}}). \tag{25}$$

Whether we make use of λ or the general quadratic formula

$$\chi^2 \sim I^{ij}\frac{\partial L}{\partial\theta_i}\frac{\partial L}{\partial\theta_j}, \tag{26}$$

where in (26) I^{ij} denotes the inverse of I_{ij} (the summation convention is used throughout this section), will thus to this order be a matter of convenience. Formula (25) is formally simpler than (26), but involves finding the simultaneous maximum-likelihood estimates $\hat{\theta}_i$.

We have (expanding $L(\hat{\theta}_i)$ about $L(\theta_i)$, rather than, as is perhaps more usual, $L(\theta_i)$ about $L(\hat{\theta}_i)$)

$$L(\hat{\theta}_i) = L(\theta_i) + \left\{(\hat{\theta}_i-\theta_i)\frac{\partial L}{\partial\theta_i} + \tfrac{1}{2}(\hat{\theta}_i-\theta_i)(\hat{\theta}_j-\theta_j)\frac{\partial^2 L}{\partial\theta_i\partial\theta_j}\right\}$$
$$+\tfrac{1}{6}(\hat{\theta}_i-\theta_i)(\hat{\theta}_j-\theta_j)(\hat{\theta}_k-\theta_k)\frac{\partial^3 L}{\partial\theta_i\partial\theta_j\partial\theta_k}+\dots. \tag{27}$$

The relation between $\hat{\theta}_i$ and L may be developed analogously to the case of one unknown quoted in §4 (equation (13)). We have

$$\begin{aligned}0 &= \frac{\partial L}{\partial\theta_i} + (\hat{\theta}_j-\theta_j)\frac{\partial^2 L}{\partial\theta_i\partial\theta_j} + \tfrac{1}{2}(\hat{\theta}_j-\theta_j)(\hat{\theta}_\kappa-\theta_\kappa)\frac{\partial^3 L}{\partial\theta_i\partial\theta_j\partial\theta_\kappa}+\dots\\ &= \frac{\partial L}{\partial\theta_i} - (\hat{\theta}_j-\theta_j)I_{ij} + \left\{(\hat{\theta}_j-\theta_j)\left(\frac{\partial^2 L}{\partial\theta_i\partial\theta_j}+I_{ij}\right) + \tfrac{1}{2}(\hat{\theta}_j-\theta_j)(\hat{\theta}_k-\theta_k)\frac{\partial^3 L}{\partial\theta_i\partial\theta_j\partial\theta_k}\right\}+\dots.\end{aligned}$$

Denote the inverse of I_{ij} by I^{ij}. Then

$$\begin{aligned}(\hat{\theta}_k-\theta_k) &= I^{ik}\frac{\partial L}{\partial\theta_i} + \left\{\frac{\partial L}{\partial\theta_h}\left[\frac{\partial^2 L}{\partial\theta_i\partial\theta_j}+I_{ij}\right]I^{hj}I^{ik} + \tfrac{1}{2}I^{ik}I^{hj}I^{gm}\frac{\partial L}{\partial\theta_h}\frac{\partial L}{\partial\theta_g}\frac{\partial^3 L}{\partial\theta_i\partial\theta_j\partial\theta_m}\right\}+\dots\\ &= I^{ik}\frac{\partial L}{\partial\theta_i}+f_k,\end{aligned} \tag{28}$$

say, where

$$f_k = \left\{\frac{\partial L}{\partial\theta_h}\left[\frac{\partial^2 L}{\partial\theta_i\partial\theta_j}I^{hj}I^{ik}+I^{hk}\right] + \tfrac{1}{2}I^{ik}I^{hj}I^{gm}\frac{\partial L}{\partial\theta_h}\frac{\partial L}{\partial\theta_g}\frac{\partial^3 L}{\partial\theta_i\partial\theta_j\partial\theta_m}\right\}+\dots.$$

---

* For a discussion of the large-sample χ^2 approximation for the likelihood ratio λ (including the more general case where L is replaced by its maximum $L'_{\text{max.}}$ with respect to a smaller subclass of the unknown parameters), see Wilks (1938), where it is shown that $-2\log\lambda$ is equivalent to a χ^2 except for terms of $O(1/\sqrt{n})$.

Hence

$$-2(L-L_{\text{max.}}) = I^{ik}\frac{\partial L}{\partial\theta_k}\frac{\partial L}{\partial\theta_i}+\left\{2f_i\frac{\partial L}{\partial\theta_i}+I^{ki}\frac{\partial L}{\partial\theta_k}I^{hj}\frac{\partial L}{\partial\theta_h}\left(\frac{\partial^2 L}{\partial\theta_i\partial\theta_j}+I_{ij}\right)\right.$$
$$\left.+2I^{ki}\frac{\partial L}{\partial\theta_k}f_i\frac{\partial^2 L}{\partial\theta_i\partial\theta_j}+\tfrac{1}{3}I^{ki}\frac{\partial L}{\partial\theta_k}I^{hj}\frac{\partial L}{\partial\theta_h}I^{gm}\frac{\partial L}{\partial\theta_g}\frac{\partial^3 L}{\partial\theta_i\partial\theta_j\partial\theta_m}\right\}+\ldots,$$

or if we put $\partial^2L/\partial\theta_i\partial\theta_j \sim -I_{ij}$ in the fourth term on the right, this may be written

$$I^{ik}\frac{\partial L}{\partial\theta_k}\frac{\partial L}{\partial\theta_i}+\left\{\frac{\partial L}{\partial\theta_i}\frac{\partial L}{\partial\theta_j}\left(I^{ik}I^{jh}\frac{\partial^2 L}{\partial\theta_k\partial\theta_h}+I^{ij}\right)+\frac{1}{3}\frac{\partial L}{\partial\theta_i}\frac{\partial L}{\partial\theta_j}\frac{\partial L}{\partial\theta_k}\frac{\partial^3 L}{\partial\theta_g\partial\theta_h\partial\theta_m}I^{ig}I^{jh}I^{km}\right\}+\ldots$$
$$= I^{ik}\frac{\partial L}{\partial\theta_k}\frac{\partial L}{\partial\theta_i}+R, \tag{29}$$

say. Now as the skewness of $\partial L/\partial\theta_k$ does not affect the distribution of the quadratic first term, it is, even to $O(1/\sqrt{n})$, a valid χ^2 quantity, as we saw in §3. It remains to consider the effect of the remainder term R. Now although R is of $O(1/\sqrt{n})$, we can readily show that $E\{R\}$, and $E\{I^{ik}\partial L/\partial\theta_k\partial L/\partial\theta_i R\}$ are of $O(1/n)$, so that $-2(L-L_{\text{max.}})$ has the same mean and variance as $I^{ik}\partial L/\partial\theta_k\partial L/\partial\theta_i$ to $O(1/\sqrt{n})$. The same method can be applied also for the higher moments. To this order the same χ^2 approximation may thus be used. Although these average values involve higher-order moments of L derivatives than we have listed, for our immediate purpose we do not require their explicit values. We merely note that

$$E\left\{\frac{\partial L}{\partial\theta_i}\frac{\partial L}{\partial\theta_j}\left(\frac{\partial^2 L}{\partial\theta_k\partial\theta_h}+I_{kh}\right)\right\}\leqslant O(n),$$
$$E\left\{\frac{\partial L}{\partial\theta_i}\frac{\partial L}{\partial\theta_j}\frac{\partial L}{\partial\theta_k}\frac{\partial^3 L}{\partial\theta_g\partial\theta_h\partial\theta_m}\right\}\leqslant O(n^2),$$
$$E\left\{\left(\frac{\partial L}{\partial\theta_i}\frac{\partial L}{\partial\theta_j}\right)\frac{\partial L}{\partial\theta_g}\frac{\partial L}{\partial\theta_m}\left(\frac{\partial^2 L}{\partial\theta_k\partial\theta_h}+I_{kh}\right)\right\}\leqslant O(n^2),$$
$$E\left\{\left(\frac{\partial L}{\partial\theta_i}\frac{\partial L}{\partial\theta_j}\right)\frac{\partial L}{\partial\theta_g}\frac{\partial L}{\partial\theta_m}\frac{\partial L}{\partial\theta_p}\frac{\partial^3 L}{\partial\theta_k\partial\theta_h\partial\theta_q}\right\}\leqslant O(n^3),$$

and as I^{ij} is $O(n^{-1})$, the result follows. It is obvious from particular examples that $-2(L-L_{\text{max.}})$ *does* require modification to $O(1/n)$ (so also does the alternative quadratic expression). In principle a general correction to $O(1/n)$ could be worked out by the above methods, though we have seen that corrections at this order are rather formidable to evaluate by the present general approach. Particular problems, e.g. in time-series analysis, might be somewhat more tractable if handled individually.

By an obvious extension of the argument, the criterion $-2(L'_{\text{max.}}-L_{\text{max.}})$, where $L'_{\text{max.}}$ refers to a smaller subclass of the unknown parameters than $L_{\text{max.}}$, is equivalent in distribution to a χ^2, also up to and including $O(1/\sqrt{n})$.

I am indebted to Mrs A. Linnert for assistance with the computations for Table 1.

REFERENCES

BARNARD, G. A. (1950). On the Fisher-Behrens test. *Biometrika*, **37**, 203.

BARTLETT, M. S. (1936). Statistical information and properties of sufficiency. *Proc. Roy. Soc.* A, **154**, 124.

BARTLETT, M. S. (1952). The statistical significance of odd bits of information. *Biometrika*, **39**, 228.

BARTLETT, M. S. (1953). Approximate confidence intervals. *Biometrika*, **40**, 12.*

CORNISH, E. A. & FISHER, R. A. (1937). Moments and cumulants in the specification of distributions. *Rev. Inst. Int. Statist.* **5**, 307.

FISHER, R. A. (1935). The fiducial argument in statistical inference. *Ann. Eugen., Lond.*, **6**, 391.

FISHER, R. A. (1941). The asymptotic approach to Behrens's integral, with further tables for the *d* test of significance. *Ann. Eugen., Lond.*, **11**, 141.

WELCH, B. L. (1947). The generalization of 'Student's' problem when several different population variances are involved. *Biometrika*, **34**, 28.

WILKS, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.* **9**, 60.

* (*Added in proof.*) Professor J. W. Tukey has just sent me a copy of his unpublished (1949) *Memorandum Report* No. 26, *Statistical Research Group, Princeton University*, on 'Standard confidence points', in which similar proposals for a single unknown parameter were put forward.

NON-NORMALITY AND TESTS ON VARIANCES

By G. E. P. BOX

Imperial Chemical Industries, Dyestuffs Division Headquarters, Blackley, Manchester, and University of North Carolina

1. Introduction

For many experimenters the most commonly used statistical tests are those for *comparing* sample means and sample variances. The test used for the equality of the means of k groups of observations is usually the analysis of variance test (or equivalently the t test when $k = 2$), whilst the test used for the equality of variances of k groups of observations is Bartlett's (1937) modification of the Neyman-Pearson (1931) L_1 test (or the equivalent F test when $k = 2$).

The tests mentioned are derived on a number of assumptions, in particular, that the observations are normally distributed. Usually, however, since little is known of the populations from which the samples are drawn, these tests are used, of necessity, as if the assumption of normality could be ignored. So far as comparative tests on means are concerned it appears (perhaps rather surprisingly) that this practice is largely justifiable, for thanks to the work of Pearson (1931), Bartlett (1935), Geary (1947), Gayen (1950 a, b), David & Johnson (1951 a, b) there is abundant evidence that these *comparative* tests on means are remarkably insensitive to general* non-normality of the parent population.

It would appear, however, that this remarkable property of 'robustness' to non-normality which these tests for comparing means possess, and without which they would be much less appropriate to the needs of the experimenter, is not necessarily shared by other statistical tests, and in particular is not shared by the tests for equality of variances, mentioned above.

The sensitivity to non-normality of the tests for comparing *two* variances was first pointed out by E. S. Pearson (1931) whose findings were confirmed by Geary (1947), Finch (1950) and Gayen (1950a). These authors showed that this test is particularly sensitive to changes in γ_2 from the normal theory value of zero. (The notation γ_1, γ_2, etc., will be used for the standardized parent cumulants. In Karl Pearson's notation $\gamma_1 = \sqrt{\beta_1}$, $\gamma_2 = \beta_2 - 3$.) In the present paper it is shown that the sensitivity is even greater when the number of variances to be compared exceeds two, and, indeed, that the sensitivity to non-normality of the L_1 criterion or the equivalent Bartlett test can be of the same order of magnitude as the sensitivity of criteria such as b_2 specifically designed to *test* normality.

It is further shown that the difference in sensitivity of the test on means on the one hand and the test on variances on the other arises from an essential difference in the nature of these two types of tests and this suggests how the extreme sensitivity to non-normality of the variance tests can be remedied.

---

* By 'general' parent non-normality is meant that the departure from normality, in particular skewness, is the same in the different groups, as could usually be assumed when the data were from an experiment in which the groups corresponded with different applied treatments to be compared. In tests in which sample means are *compared*, general skewness tends to be cancelled out; larger effects are found, however, if the skewness is in different directions in the different groups.

2. Distribution of variance test for large samples

Suppose there are k groups of independently distributed observations with n_t in the tth group and $\sum_t n_t = N$. The ith observation of the tth group is denoted by y_{ti} and the usual estimate of variance in the tth group, having $\phi_t = n_t - 1$ degrees of freedom, by s_t^2. The average of the estimated variances is denoted by s^2, where $s^2 = \sum_t \phi_t s_t^2/\Phi$ and Φ is the total number of degrees of freedom $\Sigma\phi_t = \Phi$.

The L_1 criterion due to Neyman & Pearson (1931) tests the hypothesis that the *variances* in the groups are all equal, given that the observations are normally distributed. Alternatively, M_1, the logarithmic modification due to Bartlett (1937), may be employed:

$$M_1 = \Phi \ln s^2 - \sum_t \phi_t \ln s_t^2. \tag{1}$$

Following Neyman & Pearson (1931) it may be shown that when the appropriate null hypothesis is true and on the assumption of parent normality, M_1 is distributed in large samples as χ^2_{k-1} (where χ^2_{k-1} means χ^2 with $k-1$ degrees of freedom). This is the basis of the test proposed by Bartlett (1937). He shows that, even for fairly small samples, to a close approximation M_1 is distributed as $(1+A)\chi^2_{k-1}$, where A, an adjustable constant which tends to zero for large values of ϕ_t, is given by

$$A = \frac{1}{3(k-1)}\left\{\sum_t \frac{1}{\phi_t} - \frac{1}{\Phi}\right\}. \tag{2}$$

It is convenient to define the quantity

$$d = \left(1 + \frac{1}{2}\frac{\phi}{\phi+1}\gamma_2\right)^{-1}. \tag{3}$$

When ϕ tends to infinity this becomes

$$\delta = (1 + \tfrac{1}{2}\gamma_2)^{-1}. \tag{4}$$

Thus δ is an alternative measure of kurtosis and the quantities β_2, γ_2 and δ are related as follows:

β_2	1	2	3	6	∞
γ_2	-2	-1	0	3	∞
δ	∞	2	1	0·4	0

Denote by $s_\phi^2(\gamma_2)$ an estimate of variance having ϕ degrees of freedom and based on observations drawn from a population with finite cumulants whose measure of kurtosis is γ_2. Then $s_\phi^2(0)$ is an estimate from a normal population. The estimate $s_\phi^2(\gamma_2)$ has mean σ^2 and variance $2\sigma^4/(d\phi)$. Also from the general expressions for the higher cumulants of such an estimate it is seen that as ϕ tends to infinity the distribution of $s_\phi^2(\gamma_2)$ tends to normality, and for sufficiently large values of ϕ, $s_\phi^2(\gamma_2)$ is distributed like $s_{d\phi}^2(0)$.

Denote by $M_1(0; \phi_1\phi_2, \ldots, \phi_k)$ the criterion calculated from normally distributed observations and by $M_1(\gamma_{21}\gamma_{22}\ldots\gamma_{2k}; \phi_1\phi_2\ldots\phi_k)$ the criterion calculated from k groups of observations drawn from populations with measures of kurtosis $\gamma_{21}, \gamma_{22}, \ldots, \gamma_{2t}\ldots\gamma_{2k}$. Then for large samples

$$M_1(\gamma_{21}\gamma_{22}\ldots\gamma_{2k}; \phi_1\phi_2\ldots\phi_k)$$

is distributed as

$$\Phi \ln\{\Sigma \phi_t s^2_{d_t\phi_t}(0)/\Phi\} - \sum_t \phi_t \ln s^2_{d_t\phi_t}(0). \tag{5}$$

When the samples are sufficiently large we may write δ_t for d_t, and in the important case where the kurtosis in each of the sampled populations is the same, so that $\delta_t = \delta$ for all t, $M_1(\gamma_2; \phi_1\phi_2 \ldots \phi_k)$ is distributed as

$$\delta^{-1}\{\delta\Phi \ln s^2_{\delta\Phi}(0) - \sum_t \delta\phi_t \ln s^2_{\delta\phi_t}(0)\} = \delta^{-1}M_1(0; \delta\phi_1, \delta\phi_2, \ldots, \delta\phi_k). \tag{6}$$

Consequently $M_1(\gamma_2)$ is distributed asymptotically not as χ^2_{k-1} but as $(1+\frac{1}{2}\gamma_2)\,\chi^2_{k-1}$ for any parent distribution having finite cumulants. We see therefore that M_1 is asymptotically biased if γ_2 is not zero. In particular, in large samples the mean of the distribution curve of M_1 would be $(1+\frac{1}{2}\gamma_2)\,(k-1)$ instead of $k-1$ and the standard deviation $(1+\frac{1}{2}\gamma_2)\,\{2(k-1)\}^{\frac{1}{2}}$ instead of $\{2(k-1)\}^{\frac{1}{2}}$. The discrepancy in means relative to the standard deviation would thus become larger as k was increased. Consequently divergences from normal theory probabilities, known to be large when $k = 2$, would become progressively larger for $k > 2$. Table 1 shows the true probability of exceeding the normal-theory 5 % significance levels in large samples for various values of k and γ_2 when the null-hypothesis is true.

Table 1. *True percentage chance of exceeding 5 % normal-theory significance level of M_1 in large samples from non-normal populations for various values of γ_2*

γ_2	No. of groups = k					
	2	3	5	10	20	30
2	16·6	22·4	31·5	48·9	71·8	84·9
1	11·0	13·6	17·6	25·7	38·9	49·8
0	5·0	5·0	5·0	5·0	5·0	5·0
−1	0·56	0·25	0·08	0·010	0·0004	0·00001

In the case when the null-hypothesis is not true it will be seen that changes in the level of the criterion due to non-normality and due to real differences in the group variances will tend to be 'confounded'. With leptokurtic populations the criterion will tend to show differences when none exist, while with platykurtic populations real differences will tend to be masked.

The behaviour of the test for means is in sharp contrast to that for variances discussed above. The appropriate test for the hypothesis—that the *means* in the groups are all equal, given that the variances are all equal, and that the observations are normally distributed—is the 'within and between groups' analysis of variance test. In its commonest form the criterion used is

$$F = (\text{between groups mean square})/(\text{within groups mean square})$$
$$= \left\{\sum_t \frac{n_t(\bar{y}_t - \bar{y})^2}{k-1}\right\}\bigg/ s^2.$$

Alternatively, to correspond with M_1, we could employ a form obtained from the logarithm of the likelihood ratio

$$M_2 = (N-1)\ln\{1 + (k-1)\,F/(N-k)\}.$$

For large samples Neyman & Pearson (1931) showed that assuming normality a criterion equivalent to M_2 would be distributed as χ^2_{k-1}. With certain not very severe restrictions on the parent population it is possible to show that this result is true *whether the population is* *normal or not*. Thus for large samples the criterion M_2 for means tends to follow the normal-theory distribution under conditions of non-normality. This explains to some extent the insensitivity to non-normality found by Pearson and other workers mentioned above. In contrast, the criterion M_1 for variances follows a distribution directly dependent on γ_2 however large are the samples.

It is perhaps unfortunate that the test for homogeneity of means, and the test for homogeneity of variances in the particular case of two groups, can each be brought to the form of a variance ratio test. This has sometimes led to apparently contradictory statements about the sensitivity of the 'variance ratio' to non-normality. As originally pointed out by Pearson in 1931, the two criteria are really essentially different in character but follow the same distribution when the parent is normal. The marked difference in sensitivity of the criteria is well brought out by an example taken from Gayen's work (1950*a*). A test for homogeneity of *means* of five groups of five observations leads to an F test on 4 and 20 degrees of freedom; on the normal assumption this will be distributed in the same form as the F criterion used to compare the *variances* in two independent groups of 5 and 21 observations. Gayen shows, for example, that for a certain type of non-normal distribution in which $\gamma_1 = 0$ and $\gamma_2 = 2$ the true probability of exceeding the 5 % point for the test on means would be 4·5 % (a quite trivial discrepancy), whilst that for the test on variances would be 10·2 %.

3. DISTRIBUTION OF M_1 IN SMALL SAMPLES FOR CERTAIN NON-NORMAL PARENT DISTRIBUTIONS

It will now be shown that parent distributions exist for which relations similar to (6) are obtained even for small samples.

(3·1) *Population means known*

For the moment it will be assumed that the means of the k sampled populations are known. Denote by $s^2_{.n}$ an estimate of variance having n degrees of freedom calculated from the n deviations from the known group mean η. The criterion calculated from such estimates is denoted by $M_{.1}$. If the distribution of the observations is such that $s^2_{.n}(\gamma_2)$ is distributed like $s^2_{.\delta n}(0)$ for any sample size then $ns^2_{.n}(\gamma_2)$ is distributed like $\delta^{-1}\chi^2_{\delta n}\sigma^2$, which implies that $(y-\eta)^2$ is distributed like $\delta^{-1}\chi^2_{\delta}\sigma^2$. This is true if

$$p(y) = (\delta/2\sigma^2)^{\frac{1}{2}\delta}\{\Gamma(\tfrac{1}{2}\delta)\}^{-1}\,|\,y-\eta\,|^{\delta-1}\exp\{-\delta(y-\eta)^2/2\sigma^2\} \quad (-\infty<y<+\infty;\ 0<\delta<\infty). \tag{7}$$

This is a double χ distribution, that is to say, it is a symmetrical distribution of the form of two χ distributions having δ degrees of freedom 'back to back'. It has mean η, variance σ^2, $\gamma_1 = 0$ and $\gamma_2 = 2(1-\delta)/\delta$ (that is, $\delta^{-1} = 1+\frac{1}{2}\gamma_2$ as previously defined). When $\delta = 1$ the term in $|\,y-\eta\,|$ vanishes and the distribution is normal. When δ is less than 1, γ_2 is greater than zero, and a leptokurtic distribution results. When δ is greater than 1, γ_2 is less than zero and a bimodal platykurtic distribution is obtained. For such parent distributions it is easily seen by the previous argument that

$$M_{.1}(\gamma_2;\, n_1, n_2, \ldots, n_k) \text{ is distributed } \textit{exactly} \text{ as } \delta^{-1}M_{.1}(0;\, \delta n_1, \delta n_2, \ldots, \delta n_k), \tag{8}$$

where $n_1, n_2, \ldots, n_k$ refer to the sample sizes.

(3·2) *Population means unknown*

In 1931 Le Roux carried out a very painstaking research into the distribution of the variance calculated from small samples drawn from non-normal distributions of Pearson type. He studied changes in the distribution of s^2 as the parent distribution changed, by calculating its first four moments from the moments of the parent distribution, and confirmed his findings by extensive sampling experiments. One of his discoveries was that in addition to the normal distribution a series of non-normal distributions existed (which he called $D(s^2)$ III distributions) for which the first four moments of s^2 were almost exactly those of a Pearson type III curve. These $D(s^2)$ III distributions show kurtosis accompanied by marked skewness. A selection of these adapted from Le Roux's table are shown below.

$D(s^2)$ III *distributions*

γ_1^2	0·00	0·10	0·20	0·40	0·60	0·80	1·00	1·20	1·40
γ_2	0·00	−0·04	0·03	0·24	0·48	0·73	0·99	1·25	1·51

For such curves Le Roux deduced that to a close approximation $s^2_\phi(\gamma_2)$ was distributed as $s^2_{d\phi}(0)$ for any sample size $\left(\text{where as before } d^{-1} = 1+\frac{1}{2}\frac{\phi}{\phi+1}\gamma_2\right)$, a conclusion supported by his sampling experiments. We may conclude therefore that the relation (5) is approximately true for samples of any size chosen from such populations.

In particular, if the degree of kurtosis is the same in each of the populations sampled and each variance is based on the same number ϕ of degrees of freedom, then for $D(s^2)$ III populations

$$M_1(\gamma_2;\, \phi, \phi, \ldots, \phi) \text{ is distributed approximately as } d^{-1}M_1(0;\, d\phi, d\phi, \ldots, d\phi).$$

Using Bartlett's method we see that for $D(s^2)$ III populations therefore $M_1(\gamma_2;\, \phi_1, \phi_2, \ldots, \phi_k)$ is distributed approximately not as $(1+A)\,\chi^2_{k-1}$ but as

$$d^{-1}(1+Ad^{-1})\,\chi^2_{k-1}, \tag{9}$$

where A is the correcting factor of order ϕ^{-1} defined in (2). We note in passing that departure from normality of this form would not need to be very great or the sample size very big before the effect of the constant d^{-1} completely swamped the effect of the correction term A.

4. Two groups of observations

When $k = 2$ we may use the variance ratio F as an alternative form of the M_1 criterion. Using the same argument as before and a corresponding notation it is apparent that for samples of any size drawn from double χ distributions with measures of kurtosis γ_{21} and γ_{22},

$$F(\gamma_{21}, \gamma_{22};\, n_1, n_2) \text{ is distributed exactly as } F(0;\, \delta_1 n_1, \delta_2 n_2), \tag{10}$$

whilst for samples of any size from $D(s^2)$ III populations

$$F(\gamma_{21}, \gamma_{22};\, \phi_1, \phi_2) \text{ is distributed approximately as } F(0;\, d_1\phi_1, d_2\phi_2). \tag{11}$$

These results, which are, of course, equivalent asymptotically, would be expected to apply for *any population* when the sample sizes were sufficiently large. To provide some indication of how large the samples would have to be, comparison was made with the results given by (11) and the more exact values calculated by other authors. Gayen (1950*b*) and Finch (1950) have considered the distribution of the variance ratio or the equivalent z criterion for populations defined by the first few terms of Edgeworth's series. Both these authors note that for these populations the effect of changes in γ_1 is small but large discrepancies are produced by changes in γ_2. In Fig. 1, γ_1 is assumed zero and the continuous lines represent the probability as determined by these authors, of exceeding the normal theory 5 % point for various values of γ_2. The values of γ_2 are assumed to be the same in each of the two populations. The first graph illustrates the case $\phi_1 = 24$, $\phi_2 = 60$ considered by Finch, and the second the case $\phi_1 = 4$, $\phi_2 = 20$ considered by Gayen. The dotted lines are the values given by equation (11) above. A further comparison is supplied by Pearson's (1931) sampling experiments. Pearson drew 500 pairs of samples of 5 and 20 from six experimental populations and calculated a criterion equivalent to F; this corresponds very nearly to the case $\phi_1 = 4, \phi_2 = 20$ above (ϕ_2 is actually 19 instead of 20). In his table Pearson shows the number of pairs of samples for which the criterion exceeds the 4·86 % point (which again is very close to the 5 % point considered above). These numbers reduced to percentages are plotted as circles about the second line.

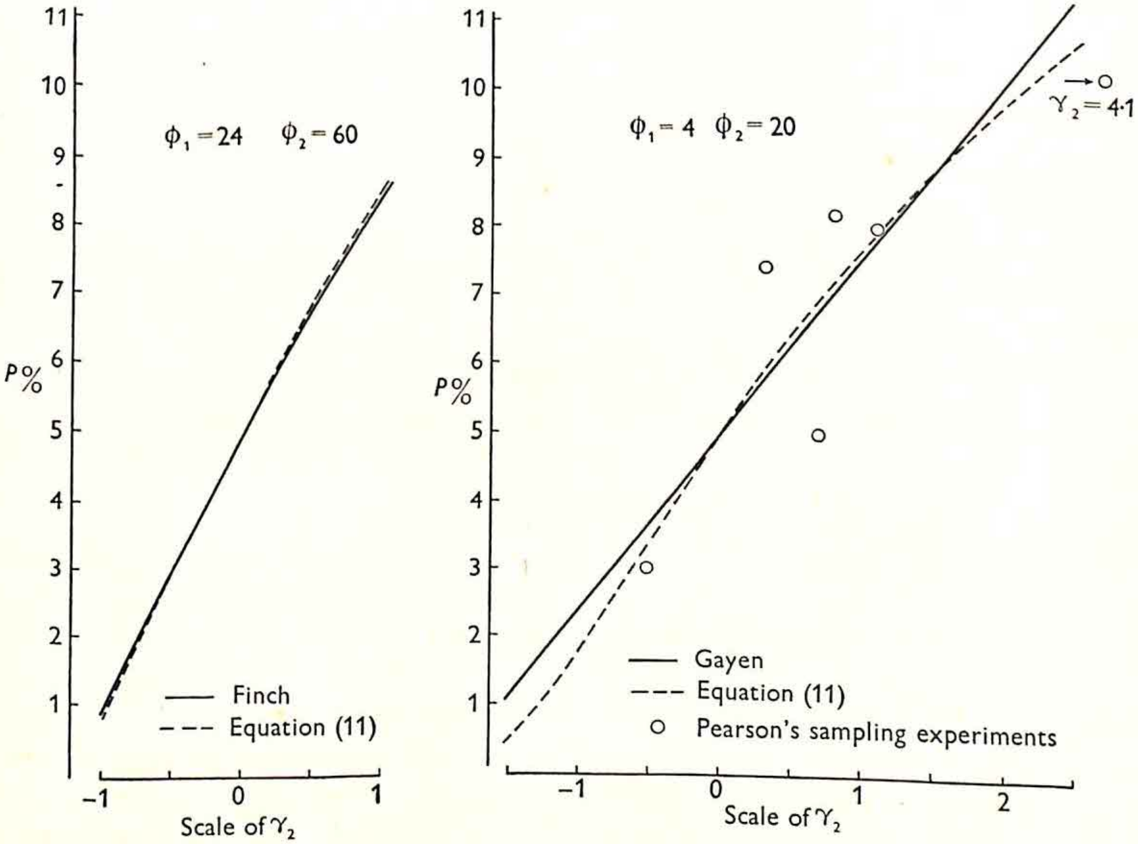


Fig. 1. Percentage probability P of exceeding normal-theory 5 % level plotted against value of γ_2.

The agreement between the simple approximation (11) and the values found by Finch is very close, as would be expected for the comparatively large sample sizes used. The extent of agreement with Gayen's results is rather surprising when the small size of one of the samples is remembered. In both approaches it is assumed, however, that the higher cumulants are finite. The marked departure from both lines which occurs when this is not so is shown by Pearson's sampling experiment from the type VII (i.e. the Student t) distribution for which $\gamma_2 = 4{\cdot}1$ and higher cumulants are infinite.

5. SAMPLING EXPERIMENTS WITH A $D(s^2)$ III DISTRIBUTION

Confirmation of (9) in the case $k = 20$ was obtained from an experiment in which samples were drawn from a population nearly of type $D(s^2)$ III form. Le Roux drew 1000 samples of $n = 20$ observations from such a distribution in which $\gamma_1^2 = 1{\cdot}0$ and $\gamma_2 = 0{\cdot}8$. (The distribution was actually a Pearson type I.) He calculated the quantity $(n-1)s^2/n$ for each of the 1000 samples and published the resulting frequency distribution. The samples were grouped but the grouping interval was fairly fine. In the present work Le Roux's observed distribution of s^2 has been resampled, the sampling being without replacement. In each experiment twenty-five groups of twenty variance estimates were reconstructed and from these, twenty-five values of M_1 were calculated. The experiment was then repeated using Le Roux's observed distribution of variance estimates for samples of $n = 5$ from the same parent distribution.

Table 2. *Characteristics of observed distributions of twenty-five values of M_1 for $k = 20$ groups, with (a) $n = 20$ observations and (b) $n = 5$ observations, drawn from a Pearson type I parent distribution for which $\gamma_1^2 = 1{\cdot}0$, $\gamma_2 = 0{\cdot}8$*

			Expected on normal theory (χ^2 approximation)	Found in samples of 25 values of M_1	Expected from equation (9) (χ^2 approximation)
$n = 20$	M_1 distributed as:		$1{\cdot}018\chi^2_{19}$		$1{\cdot}415\chi^2_{19}$
	Mean		19·4	28·6 ± 1·9*	26·9
	Variance		39·4	93·3 ± 30·3†	76·1
	No. out of 25 significant at P % point	P			
		10	2·5	13 (52 %)	10·3 (41·2 %)
		5	1·25	9 (36 %)	7·5 (30·0 %)
		1	0·25	4 (16 %)	3·2 (12·8 %)
		0·1	0·025	1 (4 %)	0·9 (3·6 %)
$n = 5$	M_1 distributed as:		$1{\cdot}088\chi^2_{19}$		$1{\cdot}473\chi^2_{19}$
	Mean		20·7	28·0 ± 1·8*	28·0
	Variance		44·9	79·1 ± 25·6†	82·4
	No. out of 25 significant at P % point	P			
		10	2·5	13 (52 %)	9·6 (38·2 %)
		5	1·25	9 (36 %)	6·8 (27·1 %)
		1	0·25	2 (8 %)	2·8 (11·0 %)
		0·1	0·025	0 (0 %)	0·7 (2·9 %)

* The quantity following the ± signs is the approximate standard error of the mean obtained from s.e.(Mean M_1) $\simeq \{V(M_1)/25\}^{\frac{1}{2}}$.

† The quantity following the ± signs is a rough estimate of the standard error of the variance of M_1 obtained by assuming that M_1 is approximately distributed in Pearson type III form and substituting $\gamma_2 = 12/(k-1)$ in the approximate formula, s.e.$\{V(M_1)\} \simeq V(M_1)\{(\gamma_2+2)/25\}^{\frac{1}{2}}$.

The two observed distributions of M_1 are shown in the sixth and seventh columns of the table in the Appendix. The main features of the observed distributions are shown in Table 2. It will be seen that the means and variances of the values of M_1 are in close agreement with the values anticipated from equation (9) and differ very markedly from those expected on normal theory. The excess of 'significant' values is also verified. For example, both for samples of $n = 20$ and $n = 5$ observations, nine out of twenty-five (or 36 %) of the values were significant at the 5 % point.

6. SAMPLING EXPERIMENTS WITH OTHER DISTRIBUTIONS

Besides the population of the $D(s^2)$ III type, Le Roux considered a variety of other populations, mostly using sampling results used by E. S. Pearson in earlier investigations, to determine empirically the distribution of the sample variance. In the present investigation Le Roux's distributions derived from the following parent distributions have been resampled in the manner already described:

Pearson type	II	III	I	IV	VII
(γ_1^2, γ_2)*	$(0{\cdot}0, -0{\cdot}5)$	$(0{\cdot}5, 0{\cdot}75)$	$(1{\cdot}0, 0{\cdot}8)$	$(1{\cdot}2, 2{\cdot}8)$	$(0{\cdot}0, 4{\cdot}1)$

From each distribution, twenty-five sets of twenty samples of s^2 have been obtained, and for these twenty-five values of M_1 calculated. These observed values of M_1 are set out in the Appendix. Table 3 shows the values of the means and variances of M_1 observed and those expected on normal theory, together with the number of values exceeding the normal theory significance points.

It is seen that large discrepancies from the normal theory distributions occur, and the discrepancies are in the directions expected. The discrepancies are larger when $n = 20$ than when $n = 5$, and for the populations studied the asymptotic result seems to give an upper limit to the discrepancy which is approached as the sample size is increased. The asymptotic result appears to be more rapidly approached when kurtosis is accompanied by skewness till for the skew $D(s^2)$ III populations, the asymptotic value is attained even for small sample sizes. The asymptotic result is approached more slowly when kurtosis is large and for the very leptokurtic type IV and type VII distributions, sample sizes greater than twenty appear to be needed before the asymptotic values are closely approached.

7. M_1 AS A TEST FOR NORMALITY

The seriousness of the discrepancies to be expected, even in small samples, may be appreciated by considering M_1 as a *test* for normality.

Suppose a sample of N observations is drawn from a population distributed in the double χ form of equation (7). We shall assume that the mean η of the population is known, so that without loss of generality we may take it to be zero. We wish to test the hypothesis H_0 that the distribution is normal, i.e. that $\delta = 1$, $\sigma > 0$, against the class of alternatives H_1 which specify that the distribution is not normal, i.e. that $\delta \neq 1$, $\sigma > 0$. Then, following through the method set out by Neyman & Pearson (1933) for the selection of best critical regions, it is easy to show that the appropriate criterion is $M_{.1} = N \ln (\Sigma y^2/N) - \Sigma \ln y^2$, and that if $M_{.1}(\alpha)$ and $M_{.1}(1-\alpha)$ are the α and $1-\alpha$ significance points for $M_{.1}$, then the inequalities

$$M_{.1} > M_{.1}(\alpha) \quad \text{and} \quad M_{.1} < M_{.1}(1-\alpha) \tag{12}$$

define a pair of common best critical regions for testing H_0 against the alternative hypotheses $\delta < 1$ $(\gamma_2 > 0)$ and $\delta > 1$ $(\gamma_2 < 0)$, respectively, where in each case σ is unknown. Thus in the particular case when there is only one observation per group, $M_{.1}$ is itself the most sensitive criterion possible for testing for a certain type of non-normality. Non-normal distributions of the class considered, which are defined by equation (7), would rarely be expected to approximate to the distribution of actual data, but the above investigation suggests that we may form some idea of the sensitivity to non-normality of $M_{.1}$ for small samples by seeing how good it is as a *test* for non-normality.

---

* The actual standardized cumulants given above differ slightly in some cases from the theoretical values due to the finite size (10,000) of the populations which Le Roux sampled.

Table 3. *Characteristics of observed distributions of twenty-five values of M_1 for $k = 20$ groups of (a) $n = 20$*, (b) $n = 5$ observations, drawn from a variety of Pearson-type parent distributions*

			Mean			Variance			No. out of 25 exceeding normal theory significance point, P			
n	γ_1^2	γ_2	Expected, normal theory	Found in samples of 25 values of M_1	Expected for $D(s^2)$ III population with same γ_2	Expected, normal theory	Found in samples of 25 values of M_1	Expected $D(s^2)$ III population with same γ_2	$P = 10\%$	$P = 5\%$	$P = 1\%$	$P = 0{\cdot}1\%$
20	0·0	−0·5	19·4	15·0	14·7	39·4	16·5	22·7	0 (0%)	0 (0%)	0 (0%)	0 (0%)
20	0·5	0·7	19·4	25·8	25·9	39·4	92·7	70·8	10 (40%)	5 (20%)	3 (12%)	2 (8%)
20	1·0	0·8	19·4	28·6	26·9	39·4	93·3	76·1	13 (52%)	9 (36%)	4 (16%)	1 (4%)
25	1·2	2·8	19·4	37·4	46·5	39·4	159·0	227·2	21 (84%)	16 (64%)	10 (40%)	5 (20%)
20	0·0	4·1	19·4	39·4	59·1	39·4	329·4	367·1	19 (76%)	17 (68%)	11 (44%)	7 (28%)
5	0·0	−0·5	20·7	19·3	16·2	44·9	36·8	27·8	2 (8%)	1 (4%)	0 (0%)	0 (0%)
5	0·5	0·7	20·7	23·0	27·0	44·9	29·4	76·9	3 (12%)	2 (8%)	0 (0%)	0 (0%)
5	1·0	0·8	20·7	28·0	28·0	44·9	79·0	82·4	13 (52%)	9 (36%)	2 (8%)	0 (0%)
5	1·2	2·8	20·7	26·8	47·8	44·9	67·5	240·0	11 (44%)	8 (32%)	1 (4%)	0 (0%)
5	0·0	4·1	20·7	28·5	61·8	44·9	192·1	401·4	9 (36%)	5 (20%)	1 (4%)	1 (4%)

* In the single case $\gamma_1^2 = 1{\cdot}2$, $\gamma_2 = 2{\cdot}8$ the sample n number taken by Le Roux was 25 instead of 20.

(7·1) *Comparison of $M_{.1}$ with criteria for testing normality*

The usual test criterion for kurtosis is b_2, the sample estimate of the moment coefficient β_2. Geary (1935*a*) proposed as a new test for kurtosis the criterion

$$a = \text{(sample mean deviation)}/\text{(sample standard deviation)}.$$

This criterion had the advantage that unlike b_2, the distribution when the null-hypothesis of normality was true could be found fairly readily. Further investigation was carried out by Geary (1935*b*, 1947) to discover whether the power of this test to pick out a departure from normality was comparable to that of b_2. Pearson (1935) studied this question by means of sampling experiments. One such experiment consisted of drawing ten samples each of seventy-six observations from each of three symmetrical populations:

Description	γ_1^2	γ_2^*
Rectangular	0·0	−1·2
Pearson type VII	0·0	4·1
Double exponential	0·0	2·9

* As before the actual standardized cumulants given above differ slightly from the theoretical values for the last two populations due to the finite size (10,000) of the population sampled.

Pearson used his results to show the relative frequency with which the test criteria b_2 and a detected departures from normality and also to show the correlation between the results of the two tests. He published not only his conclusions but also the actual samples which he used and I have used these again to calculate the corresponding values for $M_{.1}$. In my calculation of $M_{.1}$ I have used the population mean (zero in each case), whereas in Pearson's calculations of b_2 and a the sample mean was used. In samples of seventy-six, however, the discrepancy this introduces should not be large. A further difficulty arises due to the fact that the samples are grouped to the nearest unit, resulting in a frequency class having zero deviation from the mean which would give rise to an infinite value of $M_{.1}$. To overcome this I have spaced the values evenly in the interval 0·0–0·5. Thus if there were n values in the zero frequency class these have been assumed to be at $1/(4n)$, $3/(4n)$, ..., $(2n-1)/(4n)$. Fig. 2*a* shows the thirty values of $M_{.1}$ plotted against the corresponding values of b_2 and Fig. 2*b* shows the same values plotted against the values of a. It is seen that in each case marked correlation occurs, and if these graphs are compared with Pearson's Fig. 4 in which he plots a against b_2 it will be seen that the extent of correlation between $M_{.1}$ and b_2 and between $M_{.1}$ and a is similar to that between a and b_2. So far as is known, no tables are available for the significance points of M_1 (or for the equivalent criterion L_1) for the case of seventy-six groups with variance estimates each based on 1 degree of freedom. Also the asymptotic approximations of Bartlett (1937), Hartley (1940) and Box (1949) break down for large k and small numbers of degrees of freedom. Approximate significance points for $M_{.1}$ were therefore calculated by considering the distribution of $L_1^{-1} = \exp\{M_1/k\}$ which ranges from 1 to ∞. General expressions for the moments of L_1 were given by Neyman & Pearson (1931). From these the first four moments of L_1^{-1} were calculated and the appropriate Pearson type curve to approximate the distribution was found to be of type VI. It was assumed therefore that L_1^{-1} was approximately distributed as $cF_{\phi_1\phi_2}$, the curve starting at 1. By equating the first three moments of the two curves the values

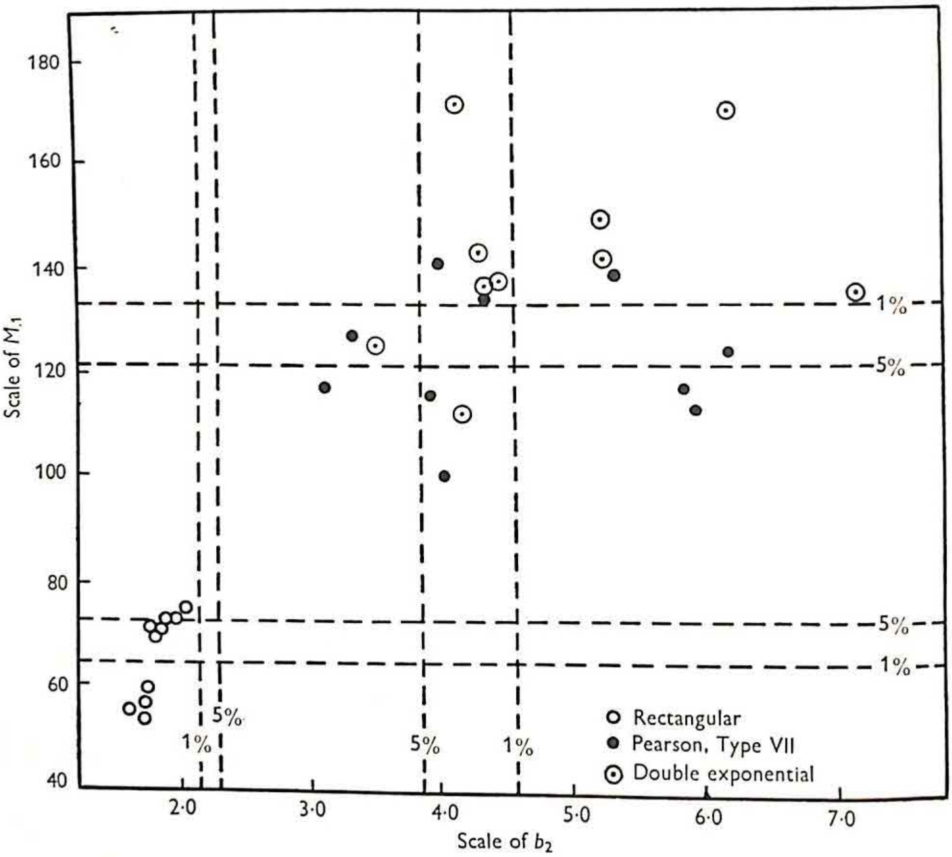


Fig. 2*a*. Comparison of $M_{.1}$ and b_2 in samples of seventy-six.

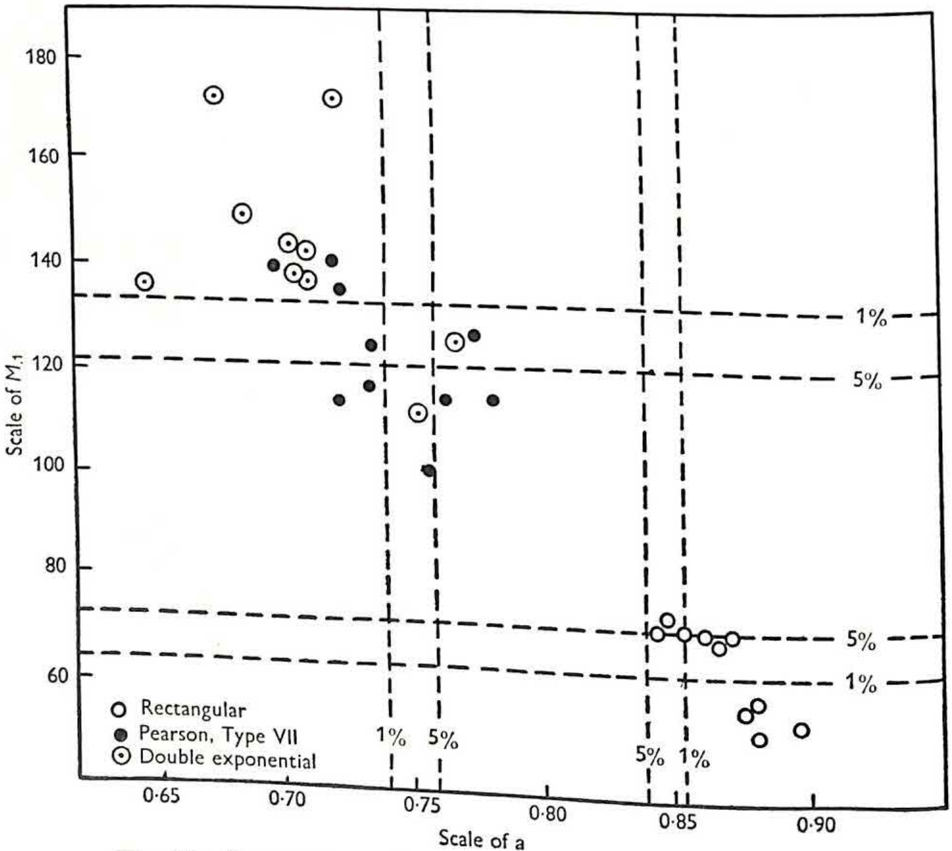


Fig. 2*b*. Comparison of $M_{.1}$ and a in samples of seventy-six.

$c = 2{\cdot}4705$, $\phi_1 = 67{\cdot}1998$, $\phi_2 = 45{\cdot}2904$ were obtained, and the approximate significance points calculated by two-way, four-point harmonic interpolation in the F tables of Merrington & Thompson (1943). The approximate significance points thus found were:

Lower		Upper	
1 %	5 %	5 %	1 %
64·1	72·3	121·2	133·3

Table 4 is an extension of Pearson's Table VII showing the number of values of b_2, a and $M_{.1}$ exceeding their 5 and 1 % significance levels.

Table 4. *Samples of seventy-six; comparison of criteria*

Distribution	Rectangular $\gamma_1^2 = 0{\cdot}0$, $\gamma_2 = -1{\cdot}2$			Pearson type VII $\gamma_1^2 = 0{\cdot}0$, $\gamma_2 = 4{\cdot}1$			Double exponential $\gamma_1^2 = 0{\cdot}0$, $\gamma_2 = 2{\cdot}9$		
	b_2	a	$M_{.1}$	b_2	a	$M_{.1}$	b_2	a	$M_{.1}$
No. of samples:									
within 5 % limits	0	0	3	2	3	5	1	1	1
between 5 and 1 % limits	0	2	3	4	1	2	5	1	1
beyond 1 % limits	10	8	4	4	6	3	4	8	8

For the rectangular and type VII distributions the criterion appears to be less sensitive to kurtosis than b_2 or a; for the double exponential distribution, however, $M_{.1}$ appears to be as sensitive as a and more sensitive than b_2. It is not of course contended that $M_{.1}$ is necessarily a practical test for kurtosis; we are concerned only to show that since b_2 and a are presumably very sensitive criteria for detecting kurtosis, $M_{.1}$ is much more sensitive *even when the sample size is small* than we should desire a test *for homogeneity of variances* to be.

In this connexion it is of some interest to reconsider some data discussed by Bartlett & Kendall (1946). The results appear as a two-way table with fifteen columns and three rows. The entries in the table are the logarithms of variance estimates s^2, each based on about 48 degrees of freedom. The authors show that by taking logarithms the data are brought to a suitable form for the application of analysis of variance. Their analysis is as follows:

	D.F.	S.S.	M.S.
Between rows	2	0·2667	0·1333
Between columns	14	0·1047	0·0075
Residual	28	0·1005	0·0036
Theoretical variance	—	—	0·0020

The first three items are calculated in the usual way, and the theoretical variance is based on the normal theory value $\kappa_2(\ln s^2) \sim 2/(\phi-1)$. Compared with this theoretical variance the residual mean square is significant ($P=0{\cdot}01$). It is deduced, therefore, that there is heterogeneity in the residual variance, and methods for further analysing this residual heterogeneity are discussed. Although familiarity with the distribution of their data (of which clearly a large quantity was available) no doubt rendered quite legitimate in this example the assumption by these authors of approximate normality, it is perhaps worth while emphasizing the dependence of this test on that assumption. For in general

$$\kappa_2(\ln s^2) \sim 2/(d\phi-1),$$

consequently a value for γ_2 of 1·57 would have made the residual and theoretical variances exactly equal. Furthermore, since any value for the theoretical variance greater than 0·0024 would fail to show the residual variance significant at the 5 % level, any hypothesis that the residual variance is homogeneous and γ_2 greater than 0·44 is not contradicted by the data at this level of significance.

8. OTHER TESTS OF VARIANCE HOMOGENEITY

The reason for the difference in sensitivity to non-normality between the criterion M_2 for testing homogeneity of means and M_1 for testing homogeneity of variances can be seen as follows. In the analysis of variance (F-form) the criterion for means may be written as P/Q, where $P = \{\Sigma n_t(\bar{y}_t-\bar{y})^2\}/(k-1)$ is the between-groups mean square and $Q = s^2$ is the within-groups mean square. When the null hypothesis is true, P and Q each provides an unbiased estimate of σ^2 whether the observations are drawn from normal populations or not, so that Q is always a standard with which P may be usefully compared. But if in M_1, the criterion for variances, we write $s_t^2 = \sigma^2(1+x_t)$ and formally expand the logarithms, we obtain $M_1 = \frac{1}{2}\sum_t \phi_t(x_t-\bar{x})^2$ plus terms in higher powers of x_t. For sufficiently large values of ϕ_t we can ignore the higher powers of x_t and we have $M_1/(k-1) \sim p/q$, where

$$p = \{\Sigma\phi_t(s_t^2-s^2)^2\}/(k-1) \quad \text{and} \quad q = 2\sigma^4.$$

Comparing these with the expressions above we see that asymptotically the M_1 test is like an analysis of variance on the sample variances instead of sample means, *but the quantity* p *corresponding to the between-groups mean square is compared not with an estimate from the internal evidence of the samples but with a theoretical value of the variance which is appropriate only when the parent distribution is normal*. In general, the variance of a variance estimate s_t^2 is $d^{-1}\times 2\sigma^4/\phi_t$; consequently the appropriate value with which to compare p would be $d^{-1}\times 2\sigma^4$ and not $2\sigma^4$.

From the above discussion it will be seen that to obtain a criterion less dependent on γ_2, information on the variation to be expected in the sample variances gathered from internal evidence in the samples should be utilized. A test of variances should be 'studentized' for the fourth moment just as a test of means is studentized for the second moment.

(8·1) *Other tests not utilizing information within groups*

Other tests of variance homogeneity which do not utilize evidence on variance variability within the samples are equally sensitive to non-normality. In fact for $D(s^2)$ III parent distributions and, asymptotically, for all distributions with finite cumulants any criterion

$f\{s^2_{\phi_1}(\gamma_{21}), s^2_{\phi_2}(\gamma_{22}), \ldots, s^2_{\phi_k}(\gamma_{2k})\}$ will be distributed like $f\{s^2_{d_1\phi_1}(0), s^2_{d_2\phi_2}(0), \ldots, s^2_{d_k\phi_k}(0)\}$. Examples of such tests are those proposed by Cochran (1941) and by Hartley (1950). Cochran's criterion is $g = s^2_{\text{max.}}/\Sigma s^2_t$, whilst that proposed by Hartley is $F_{\text{max.}} = s^2_{\text{max.}}/s^2_{\text{min.}}$, where $s^2_{\text{max.}}$ and $s^2_{\text{min.}}$ are the largest and smallest of the group variances $s^2_1, s^2_2, \ldots, s^2_t, \ldots, s^2_k$. Using the same approximation as that adopted by Hartley it is a simple matter to calculate the true chance of exceeding the normal theory significance points for a $D(s^2)$ III parent population. For instance with k groups of 21 observations drawn at random from such a population in which $\gamma_2 = 1$ the approximate percentage chance of exceeding the normal theory 5 % significance point would be:

	$k = 5$	$k = 10$	$k = 20$
M_1	17·5	25·2	37·3
$F_{\text{max.}}$	17·3	23·3	30·3

The discrepancies using $F_{\text{max.}}$ are seen to be of similar magnitude to those found with M_1, and similar results are to be expected using Cochran's g. The multivariate tests for the constancy of the variance covariance matrix from one group of observations to another (Wilks, 1932; Bishop, 1939; Box, 1949, 1950) would be expected to be equally dependent on the assumption of multivariate normality.

Sensitivity to non-normality is found also with sequential tests. For example, Wald (1947) discussed the use of the sequential likelihood ratio test of the hypothesis H_0 that $\sigma^2 = \sigma^2_0$ when the alternative H_1 is that $\sigma^2 = \sigma^2_1$ and the observations $y_1, y_2, \ldots$ were drawn from normal populations with known mean η. Suppose the observations were drawn, not from a normal universe, but from the double χ population whose distribution is given by equation (7). Then it will be found that the logarithm of the likelihood ratio $L(\gamma_2)$ is equal to $\delta L(0)$ (where $L(0)$ is the logarithm of the likelihood ratio when the parent distribution is normal). Consequently if the quantity $L(0)$ is calculated when γ_2 is not zero and referred to limits $\ln(\beta/(1-\alpha))$ and $\ln((1-\beta)/\alpha)$, this is equivalent to referring the actual likelihood $\delta L(0)$ to the limits $\delta \ln\{\beta/(1-\alpha)\}$ and $\delta \ln\{(1-\beta)/\alpha\}$. The actual risks of error of the two kinds will therefore be α' and β' chosen so that $\beta'/(1-\alpha') = \{\beta/(1-\alpha)\}^\delta$ and $(1-\beta')/\alpha' = \{(1-\beta)/\alpha\}^\delta$. For $100\alpha = 100\beta = 5$ the percentage risk of errors of each of the two kinds is given below for various values of γ_2:

γ_2	-1	0	1	2
$100\alpha' = 100\beta'$	0·28	5·00	12·32	18·66

As with the tests with fixed sample sizes the result will be approximately true for any population with finite cumulants when the average sample sizes are not small. Roughly $\alpha' = \alpha^\delta$, $\beta' = \beta^\delta$ for small values of α and β. Similar difficulties would be expected with other sequential tests relating to the true value of and equality of variances; for example, with the tests proposed by Girshick (1946) and the tests based on ranges suggested by Cox (1949).

(8·2) *A test utilizing within-group information*

A test on variances less sensitive to parent non-normality must clearly utilize the within-group information in some way. Although the author believes that a better approach is available, one immediately practical method is to split up the groups of observations into sub-groups, and carry out an analysis of variance between and within groups on the logarithms of the sub-group variances, following Bartlett & Kendall (1946), who have shown the value of the logarithmic transformation in bringing variance data to a form suitable for the application of analysis of variance. The question of what sizes should best be taken for the subgroups requires further investigation.

Table 5. *Criteria calculated for ten samples of twenty groups of twenty observations drawn from a rectangular population*

Source of sample*		Test of means	Tests of variances			
Page	Rows	M_2	M_1	M_2 on logarithms of subgroup variances		
				2 sub-groups of 10	5 sub-groups of 4	10 sub-groups of 2
I	1–20	14·5	6·7	26·2	10·3	34·0
I	21–40	22·7	8·6	18·6	16·1	18·1
I and II	41–50 1–10	13·8	10·8	29·6	29·3	19·3
II	11–30	12·6	16·7	30·7	35·0	18·1
II	31–50	19·6	9·9	40·6	20·3	10·3
III	1–20	26·9	5·2	17·9	10·8	11·6
III	21–40	12·5	16·1	40·8	29·6	19·8
III and IV	41–50 1–10	18·0	7·4	34·1	21·6	34·1
IV	11–30	23·8	10·3	37·9	28·2	19·5
IV	31–50	11·8	10·3	30·6	32·4	29·6
Mean:	Found	17·6 ± 1·7	10·2 ± 1·2	30·7 ± 2·6	23·4 ± 2·8	21·4 ± 2·7
	Expected on normal theory	19·5	19·4	27·0	22·1	21·3
Variance:	Found	29·3 ± 15·6	13·9 ± 7·2	66·1 ± 35·3	78·7 ± 42·0	71·0 ± 37·9
	Expected on normal theory	40·1	39·4	79·0	51·5	48·1

* The groups of observations consisted in each case of the first twenty columns of numbers from pages of Fisher & Yates's tables. The page numbers and the numbers of the rows are shown above.

The results of a small sampling experiment of some interest in this connexion are set out in Table 5. Ten samples of twenty groups of twenty observations were drawn from the table of random numbers prepared by Fisher & Yates (1938). The parent distribution was thus effectively rectangular. As a test of group to group homogeneity of *means* a 'between and within groups' analysis of variance was performed for each of the ten samples of twenty groups and the ten resulting values of M_2 calculated. These are shown in the third column of the table. Four different tests for homogeneity of the variances were applied to each sample. The first was the M_1 test, the results for which are shown in the fourth column of the

REFERENCES

BARNARD, G. A. (1950). On the Fisher-Behrens test. *Biometrika*, **37**, 203.

BARTLETT, M. S. (1936). Statistical information and properties of sufficiency. *Proc. Roy. Soc.* A, **154**, 124.

BARTLETT, M. S. (1952). The statistical significance of odd bits of information. *Biometrika*, **39**, 228.

BARTLETT, M. S. (1953). Approximate confidence intervals. *Biometrika*, **40**, 12.*

CORNISH, E. A. & FISHER, R. A. (1937). Moments and cumulants in the specification of distributions. *Rev. Inst. Int. Statist.* **5**, 307.

FISHER, R. A. (1935). The fiducial argument in statistical inference. *Ann. Eugen., Lond.*, **6**, 391.

FISHER, R. A. (1941). The asymptotic approach to Behrens's integral, with further tables for the *d* test of significance. *Ann. Eugen., Lond.*, **11**, 141.

WELCH, B. L. (1947). The generalization of 'Student's' problem when several different population variances are involved. *Biometrika*, **34**, 28.

WILKS, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.* **9**, 60.

* (*Added in proof.*) Professor J. W. Tukey has just sent me a copy of his unpublished (1949) *Memorandum Report* No. 26, *Statistical Research Group, Princeton University*, on 'Standard confidence points', in which similar proposals for a single unknown parameter were put forward.

NON-NORMALITY AND TESTS ON VARIANCES

By G. E. P. BOX

Imperial Chemical Industries, Dyestuffs Division Headquarters, Blackley, Manchester, and University of North Carolina

1. Introduction

For many experimenters the most commonly used statistical tests are those for *comparing* sample means and sample variances. The test used for the equality of the means of k groups of observations is usually the analysis of variance test (or equivalently the t test when $k = 2$), whilst the test used for the equality of variances of k groups of observations is Bartlett's (1937) modification of the Neyman-Pearson (1931) L_1 test (or the equivalent F test when $k = 2$).

The tests mentioned are derived on a number of assumptions, in particular, that the observations are normally distributed. Usually, however, since little is known of the populations from which the samples are drawn, these tests are used, of necessity, as if the assumption of normality could be ignored. So far as comparative tests on means are concerned it appears (perhaps rather surprisingly) that this practice is largely justifiable, for thanks to the work of Pearson (1931), Bartlett (1935), Geary (1947), Gayen (1950 *a*, *b*), David & Johnson (1951 *a*, *b*) there is abundant evidence that these *comparative* tests on means are remarkably insensitive to general* non-normality of the parent population.

It would appear, however, that this remarkable property of 'robustness' to non-normality which these tests for comparing means possess, and without which they would be much less appropriate to the needs of the experimenter, is not necessarily shared by other statistical tests, and in particular is not shared by the tests for equality of variances, mentioned above.

The sensitivity to non-normality of the tests for comparing *two* variances was first pointed out by E. S. Pearson (1931) whose findings were confirmed by Geary (1947), Finch (1950) and Gayen (1950 *a*). These authors showed that this test is particularly sensitive to changes in γ_2 from the normal theory value of zero. (The notation γ_1, γ_2, etc., will be used for the standardized parent cumulants. In Karl Pearson's notation $\gamma_1 = \surd\beta_1$, $\gamma_2 = \beta_2 - 3$.) In the present paper it is shown that the sensitivity is even greater when the number of variances to be compared exceeds two, and, indeed, that the sensitivity to non-normality of the L_1 criterion or the equivalent Bartlett test can be of the same order of magnitude as the sensitivity of criteria such as b_2 specifically designed to *test* normality.

It is further shown that the difference in sensitivity of the test on means on the one hand and the test on variances on the other arises from an essential difference in the nature of these two types of tests and this suggests how the extreme sensitivity to non-normality of the variance tests can be remedied.

---

* By 'general' parent non-normality is meant that the departure from normality, in particular skewness, is the same in the different groups, as could usually be assumed when the data were from an experiment in which the groups corresponded with different applied treatments to be compared. In tests in which sample means are *compared*, general skewness tends to be cancelled out; larger effects are found, however, if the skewness is in different directions in the different groups.

2. Distribution of variance test for large samples

Suppose there are k groups of independently distributed observations with n_t in the tth group and $\sum_t n_t = N$. The ith observation of the tth group is denoted by y_{ti} and the usual estimate of variance in the tth group, having $\phi_t = n_t - 1$ degrees of freedom, by s_t^2. The average of the estimated variances is denoted by s^2, where $s^2 = \sum_t \phi_t s_t^2/\Phi$ and Φ is the total number of degrees of freedom $\Sigma\phi_t = \Phi$.

The L_1 criterion due to Neyman & Pearson (1931) tests the hypothesis that the *variances* in the groups are all equal, given that the observations are normally distributed. Alternatively, M_1, the logarithmic modification due to Bartlett (1937), may be employed:

$$M_1 = \Phi \ln s^2 - \sum_t \phi_t \ln s_t^2. \tag{1}$$

Following Neyman & Pearson (1931) it may be shown that when the appropriate null hypothesis is true and on the assumption of parent normality, M_1 is distributed in large samples as χ^2_{k-1} (where χ^2_{k-1} means χ^2 with $k-1$ degrees of freedom). This is the basis of the test proposed by Bartlett (1937). He shows that, even for fairly small samples, to a close approximation M_1 is distributed as $(1+A)\chi^2_{k-1}$, where A, an adjustable constant which tends to zero for large values of ϕ_t, is given by

$$A = \frac{1}{3(k-1)}\left\{\sum_t \frac{1}{\phi_t} - \frac{1}{\Phi}\right\}. \tag{2}$$

It is convenient to define the quantity

$$d = \left(1 + \frac{1}{2}\frac{\phi}{\phi+1}\gamma_2\right)^{-1}. \tag{3}$$

When ϕ tends to infinity this becomes

$$\delta = (1+\tfrac{1}{2}\gamma_2)^{-1}. \tag{4}$$

Thus δ is an alternative measure of kurtosis and the quantities β_2, γ_2 and δ are related as follows:

β_2	1	2	3	6	∞
γ_2	-2	-1	0	3	∞
δ	∞	2	1	0·4	0

Denote by $s_\phi^2(\gamma_2)$ an estimate of variance having ϕ degrees of freedom and based on observations drawn from a population with finite cumulants whose measure of kurtosis is γ_2. Then $s_\phi^2(0)$ is an estimate from a normal population. The estimate $s_\phi^2(\gamma_2)$ has mean σ^2 and variance $2\sigma^4/(d\phi)$. Also from the general expressions for the higher cumulants of such an estimate it is seen that as ϕ tends to infinity the distribution of $s_\phi^2(\gamma_2)$ tends to normality, and for sufficiently large values of ϕ, $s_\phi^2(\gamma_2)$ is distributed like $s_{d\phi}^2(0)$.

Denote by $M_1(0;\, \phi_1\phi_2, \ldots, \phi_k)$ the criterion calculated from normally distributed observations and by $M_1(\gamma_{21}\gamma_{22}\cdots\gamma_{2k};\, \phi_1\phi_2\cdots\phi_k)$ the criterion calculated from k groups of observations drawn from populations with measures of kurtosis $\gamma_{21}, \gamma_{22}, \ldots, \gamma_{2t}\cdots\gamma_{2k}$. Then for large samples

$$M_1(\gamma_{21}\gamma_{22}\cdots\gamma_{2k};\, \phi_1\phi_2\cdots\phi_k)$$

is distributed as

$$\Phi \ln\{\Sigma\, \phi_t s^2_{d_t\phi_t}(0)/\Phi\} - \sum_t \phi_t \ln s^2_{d_t\phi_t}(0). \tag{5}$$

When the samples are sufficiently large we may write δ_t for d_t, and in the important case where the kurtosis in each of the sampled populations is the same, so that $\delta_t = \delta$ for all t, $M_1(\gamma_2; \phi_1 \phi_2 \ldots \phi_k)$ is distributed as

$$\delta^{-1}\{\delta\Phi \ln s^2_{\delta\Phi}(0) - \sum_t \delta\phi_t \ln s^2_{\delta\phi_t}(0)\} = \delta^{-1} M_1(0; \delta\phi_1, \delta\phi_2, \ldots, \delta\phi_k). \tag{6}$$

Consequently $M_1(\gamma_2)$ is distributed asymptotically not as χ^2_{k-1} but as $(1+\frac{1}{2}\gamma_2)\,\chi^2_{k-1}$ for any parent distribution having finite cumulants. We see therefore that M_1 is asymptotically biased if γ_2 is not zero. In particular, in large samples the mean of the distribution curve of M_1 would be $(1+\frac{1}{2}\gamma_2)(k-1)$ instead of $k-1$ and the standard deviation $(1+\frac{1}{2}\gamma_2)\{2(k-1)\}^{\frac{1}{2}}$ instead of $\{2(k-1)\}^{\frac{1}{2}}$. The discrepancy in means relative to the standard deviation would thus become larger as k was increased. Consequently divergences from normal theory probabilities, known to be large when $k = 2$, would become progressively larger for $k > 2$. Table 1 shows the true probability of exceeding the normal-theory 5 % significance levels in large samples for various values of k and γ_2 when the null-hypothesis is true.

Table 1. *True percentage chance of exceeding 5 % normal-theory significance level of M_1 in large samples from non-normal populations for various values of γ_2*

γ_2	No. of groups = k					
	2	3	5	10	20	30
2	16·6	22·4	31·5	48·9	71·8	84·9
1	11·0	13·6	17·6	25·7	38·9	49·8
0	5·0	5·0	5·0	5·0	5·0	5·0
−1	0·56	0·25	0·08	0·010	0·0004	0·00001

In the case when the null-hypothesis is not true it will be seen that changes in the level of the criterion due to non-normality and due to real differences in the group variances will tend to be 'confounded'. With leptokurtic populations the criterion will tend to show differences when none exist, while with platykurtic populations real differences will tend to be masked.

The behaviour of the test for means is in sharp contrast to that for variances discussed above. The appropriate test for the hypothesis—that the *means* in the groups are all equal, given that the variances are all equal, and that the observations are normally distributed—is the 'within and between groups' analysis of variance test. In its commonest form the criterion used is

$$F = \text{(between groups mean square)}/\text{(within groups mean square)}$$
$$= \left\{\sum_t \frac{n_t(\bar{y}_t - \bar{y})^2}{k-1}\right\} \Big/ s^2.$$

Alternatively, to correspond with M_1, we could employ a form obtained from the logarithm of the likelihood ratio

$$M_2 = (N-1)\ln\{1+(k-1)\,F/(N-k)\}.$$

For large samples Neyman & Pearson (1931) showed that assuming normality a criterion equivalent to M_2 would be distributed as χ^2_{k-1}. With certain not very severe restrictions on the parent population it is possible to show that this result is true *whether the population is* *normal or not*. Thus for large samples the criterion M_2 for means tends to follow the normal-theory distribution under conditions of non-normality. This explains to some extent the insensitivity to non-normality found by Pearson and other workers mentioned above. In contrast, the criterion M_1 for variances follows a distribution directly dependent on γ_2 however large are the samples.

It is perhaps unfortunate that the test for homogeneity of means, and the test for homogeneity of variances in the particular case of two groups, can each be brought to the form of a variance ratio test. This has sometimes led to apparently contradictory statements about the sensitivity of the 'variance ratio' to non-normality. As originally pointed out by Pearson in 1931, the two criteria are really essentially different in character but follow the same distribution when the parent is normal. The marked difference in sensitivity of the criteria is well brought out by an example taken from Gayen's work (1950*a*). A test for homogeneity of *means* of five groups of five observations leads to an F test on 4 and 20 degrees of freedom; on the normal assumption this will be distributed in the same form as the F criterion used to compare the *variances* in two independent groups of 5 and 21 observations. Gayen shows, for example, that for a certain type of non-normal distribution in which $\gamma_1 = 0$ and $\gamma_2 = 2$ the true probability of exceeding the 5 % point for the test on means would be 4·5 % (a quite trivial discrepancy), whilst that for the test on variances would be 10·2 %.

3. DISTRIBUTION OF M_1 IN SMALL SAMPLES FOR CERTAIN NON-NORMAL PARENT DISTRIBUTIONS

It will now be shown that parent distributions exist for which relations similar to (6) are obtained even for small samples.

(3·1) *Population means known*

For the moment it will be assumed that the means of the k sampled populations are known. Denote by $s^2_{.n}$ an estimate of variance having n degrees of freedom calculated from the n deviations from the known group mean η. The criterion calculated from such estimates is denoted by $M_{.1}$. If the distribution of the observations is such that $s^2_{.n}(\gamma_2)$ is distributed like $s^2_{.\delta n}(0)$ for any sample size then $ns^2_{.n}(\gamma_2)$ is distributed like $\delta^{-1}\chi^2_{\delta n}\sigma^2$, which implies that $(y-\eta)^2$ is distributed like $\delta^{-1}\chi^2_\delta\sigma^2$. This is true if

$$p(y) = (\delta/2\sigma^2)^{\frac{1}{2}\delta}\{\Gamma(\tfrac{1}{2}\delta)\}^{-1}\,|\,y-\eta\,|^{\delta-1}\exp\{-\delta(y-\eta)^2/2\sigma^2\} \quad (-\infty<y<+\infty;\ 0<\delta<\infty). \tag{7}$$

This is a double χ distribution, that is to say, it is a symmetrical distribution of the form of two χ distributions having δ degrees of freedom 'back to back'. It has mean η, variance σ^2, $\gamma_1 = 0$ and $\gamma_2 = 2(1-\delta)/\delta$ (that is, $\delta^{-1} = 1+\frac{1}{2}\gamma_2$ as previously defined). When $\delta = 1$ the term in $|\,y-\eta\,|$ vanishes and the distribution is normal. When δ is less than 1, γ_2 is greater than zero, and a leptokurtic distribution results. When δ is greater than 1, γ_2 is less than zero and a bimodal platykurtic distribution is obtained. For such parent distributions it is easily seen by the previous argument that

$$M_{.1}(\gamma_2;\, n_1, n_2, \ldots, n_k) \text{ is distributed } \textit{exactly} \text{ as } \delta^{-1}M_{.1}(0;\, \delta n_1, \delta n_2, \ldots, \delta n_k), \tag{8}$$

where $n_1, n_2, \ldots, n_k$ refer to the sample sizes.

(3·2) *Population means unknown*

In 1931 Le Roux carried out a very painstaking research into the distribution of the variance calculated from small samples drawn from non-normal distributions of Pearson type. He studied changes in the distribution of s^2 as the parent distribution changed, by calculating its first four moments from the moments of the parent distribution, and confirmed his findings by extensive sampling experiments. One of his discoveries was that in addition to the normal distribution a series of non-normal distributions existed (which he called $D(s^2)$ III distributions) for which the first four moments of s^2 were almost exactly those of a Pearson type III curve. These $D(s^2)$ III distributions show kurtosis accompanied by marked skewness. A selection of these adapted from Le Roux's table are shown below.

$D(s^2)$ III *distributions*

γ_1^2	0·00	0·10	0·20	0·40	0·60	0·80	1·00	1·20	1·40
γ_2	0·00	−0·04	0·03	0·24	0·48	0·73	0·99	1·25	1·51

For such curves Le Roux deduced that to a close approximation $s^2_\phi(\gamma_2)$ was distributed as $s^2_{d\phi}(0)$ for any sample size $\left(\text{where as before } d^{-1} = 1 + \frac{1}{2}\frac{\phi}{\phi+1}\gamma_2\right)$, a conclusion supported by his sampling experiments. We may conclude therefore that the relation (5) is approximately true for samples of any size chosen from such populations.

In particular, if the degree of kurtosis is the same in each of the populations sampled and each variance is based on the same number ϕ of degrees of freedom, then for $D(s^2)$ III populations

$$M_1(\gamma_2;\, \phi, \phi, \ldots, \phi) \text{ is distributed approximately as } d^{-1}M_1(0;\, d\phi, d\phi, \ldots, d\phi).$$

Using Bartlett's method we see that for $D(s^2)$ III populations therefore $M_1(\gamma_2;\, \phi_1, \phi_2, \ldots, \phi_k)$ is distributed approximately not as $(1+A)\,\chi^2_{k-1}$ but as

$$d^{-1}(1+Ad^{-1})\,\chi^2_{k-1}, \tag{9}$$

where A is the correcting factor of order ϕ^{-1} defined in (2). We note in passing that departure from normality of this form would not need to be very great or the sample size very big before the effect of the constant d^{-1} completely swamped the effect of the correction term A.

4. Two groups of observations

When $k = 2$ we may use the variance ratio F as an alternative form of the M_1 criterion. Using the same argument as before and a corresponding notation it is apparent that for samples of any size drawn from double χ distributions with measures of kurtosis γ_{21} and γ_{22},

$$F(\gamma_{21}, \gamma_{22};\, n_1, n_2) \text{ is distributed exactly as } F(0;\, \delta_1 n_1, \delta_2 n_2), \tag{10}$$

whilst for samples of any size from $D(s^2)$ III populations

$$F(\gamma_{21}, \gamma_{22};\, \phi_1, \phi_2) \text{ is distributed approximately as } F(0;\, d_1\phi_1, d_2\phi_2). \tag{11}$$

These results, which are, of course, equivalent asymptotically, would be expected to apply for *any population* when the sample sizes were sufficiently large. To provide some indication of how large the samples would have to be, comparison was made with the results given by (11) and the more exact values calculated by other authors. Gayen (1950*b*) and Finch (1950) have considered the distribution of the variance ratio or the equivalent z criterion for populations defined by the first few terms of Edgeworth's series. Both these authors note that for these populations the effect of changes in γ_1 is small but large discrepancies are produced by changes in γ_2. In Fig. 1, γ_1 is assumed zero and the continuous lines represent the probability as determined by these authors, of exceeding the normal theory 5 % point for various values of γ_2. The values of γ_2 are assumed to be the same in each of the two populations. The first graph illustrates the case $\phi_1 = 24$, $\phi_2 = 60$ considered by Finch, and the second the case $\phi_1 = 4$, $\phi_2 = 20$ considered by Gayen. The dotted lines are the values given by equation (11) above. A further comparison is supplied by Pearson's (1931) sampling experiments. Pearson drew 500 pairs of samples of 5 and 20 from six experimental populations and calculated a criterion equivalent to F; this corresponds very nearly to the case $\phi_1 = 4, \phi_2 = 20$ above (ϕ_2 is actually 19 instead of 20). In his table Pearson shows the number of pairs of samples for which the criterion exceeds the 4·86 % point (which again is very close to the 5 % point considered above). These numbers reduced to percentages are plotted as circles about the second line.

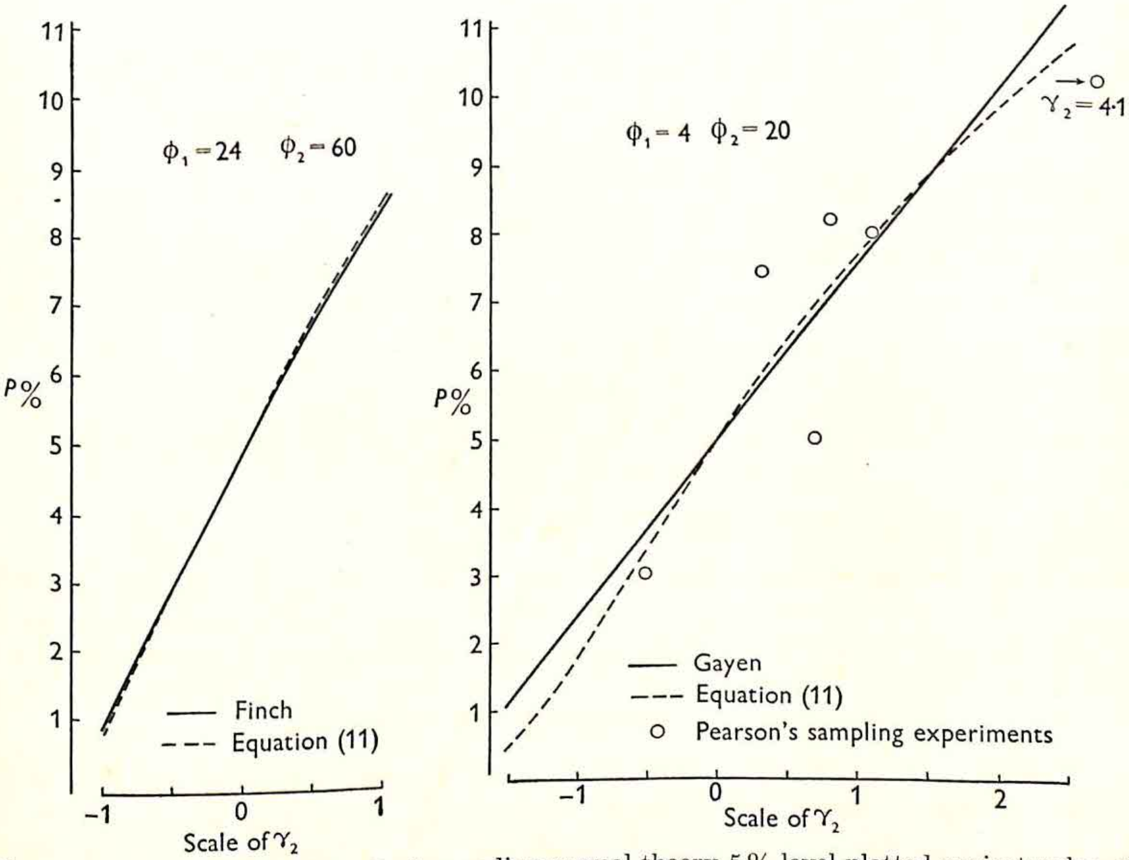


Fig. 1. Percentage probability P of exceeding normal-theory 5 % level plotted against value of γ_2.

The agreement between the simple approximation (11) and the values found by Finch is very close, as would be expected for the comparatively large sample sizes used. The extent of agreement with Gayen's results is rather surprising when the small size of one of the samples is remembered. In both approaches it is assumed, however, that the higher cumulants are finite. The marked departure from both lines which occurs when this is not so is shown by Pearson's sampling experiment from the type VII (i.e. the Student t) distribution for which $\gamma_2 = 4\cdot 1$ and higher cumulants are infinite.

5. Sampling experiments with a $D(s^2)$ III distribution

Confirmation of (9) in the case $k = 20$ was obtained from an experiment in which samples were drawn from a population nearly of type $D(s^2)$ III form. Le Roux drew 1000 samples of $n = 20$ observations from such a distribution in which $\gamma_1^2 = 1\cdot 0$ and $\gamma_2 = 0\cdot 8$. (The distribution was actually a Pearson type I.) He calculated the quantity $(n-1)\,s^2/n$ for each of the 1000 samples and published the resulting frequency distribution. The samples were grouped but the grouping interval was fairly fine. In the present work Le Roux's observed distribution of s^2 has been resampled, the sampling being without replacement. In each experiment twenty-five groups of twenty variance estimates were reconstructed and from these, twenty-five values of M_1 were calculated. The experiment was then repeated using Le Roux's observed distribution of variance estimates for samples of $n = 5$ from the same parent distribution.

Table 2. *Characteristics of observed distributions of twenty-five values of M_1 for $k = 20$ groups, with (a) $n = 20$ observations and (b) $n = 5$ observations, drawn from a Pearson type I parent distribution for which $\gamma_1^2 = 1{\cdot}0$, $\gamma_2 = 0{\cdot}8$*

			Expected on normal theory (χ^2 approximation)	Found in samples of 25 values of M_1	Expected from equation (9) (χ^2 approximation)
$n = 20$	M_1 distributed as:		$1{\cdot}018\chi^2_{19}$		$1{\cdot}415\chi^2_{19}$
	Mean		19·4	28·6 ± 1·9*	26·9
	Variance		39·4	93·3 ± 30·3†	76·1
	No. out of 25 significant at P % point	P 10	2·5	13 (52 %)	10·3 (41·2 %)
		5	1·25	9 (36 %)	7·5 (30·0 %)
		1	0·25	4 (16 %)	3·2 (12·8 %)
		0·1	0·025	1 (4 %)	0·9 (3·6 %)
$n = 5$	M_1 distributed as:		$1{\cdot}088\chi^2_{19}$		$1{\cdot}473\chi^2_{19}$
	Mean		20·7	28·0 ± 1·8*	28·0
	Variance		44·9	79·1 ± 25·6†	82·4
	No. out of 25 significant at P % point	P 10	2·5	13 (52 %)	9·6 (38·2 %)
		5	1·25	9 (36 %)	6·8 (27·1 %)
		1	0·25	2 (8 %)	2·8 (11·0 %)
		0·1	0·025	0 (0 %)	0·7 (2·9 %)

* The quantity following the ± signs is the approximate standard error of the mean obtained from s.e.(Mean M_1) $\simeq \{V(M_1)/25\}^{\frac{1}{2}}$.

† The quantity following the ± signs is a rough estimate of the standard error of the variance of M_1 obtained by assuming that M_1 is approximately distributed in Pearson type III form and substituting $\gamma_2 = 12/(k-1)$ in the approximate formula, s.e.$\{V(M_1)\} \simeq V(M_1)\{(\gamma_2+2)/25\}^{\frac{1}{2}}$.

The two observed distributions of M_1 are shown in the sixth and seventh columns of the table in the Appendix. The main features of the observed distributions are shown in Table 2. It will be seen that the means and variances of the values of M_1 are in close agreement with the values anticipated from equation (9) and differ very markedly from those expected on normal theory. The excess of 'significant' values is also verified. For example, both for samples of $n = 20$ and $n = 5$ observations, nine out of twenty-five (or 36 %) of the values were significant at the 5 % point.

6. SAMPLING EXPERIMENTS WITH OTHER DISTRIBUTIONS

Besides the population of the $D(s^2)$ III type, Le Roux considered a variety of other populations, mostly using sampling results used by E. S. Pearson in earlier investigations, to determine empirically the distribution of the sample variance. In the present investigation Le Roux's distributions derived from the following parent distributions have been resampled in the manner already described:

Pearson type	II	III	I	IV	VII
(γ_1^2, γ_2)*	$(0\cdot0, -0\cdot5)$	$(0\cdot5, 0\cdot75)$	$(1\cdot0, 0\cdot8)$	$(1\cdot2, 2\cdot8)$	$(0\cdot0, 4\cdot1)$

From each distribution, twenty-five sets of twenty samples of s^2 have been obtained, and for these twenty-five values of M_1 calculated. These observed values of M_1 are set out in the Appendix. Table 3 shows the values of the means and variances of M_1 observed and those expected on normal theory, together with the number of values exceeding the normal theory significance points.

It is seen that large discrepancies from the normal theory distributions occur, and the discrepancies are in the directions expected. The discrepancies are larger when $n = 20$ than when $n = 5$, and for the populations studied the asymptotic result seems to give an upper limit to the discrepancy which is approached as the sample size is increased. The asymptotic result appears to be more rapidly approached when kurtosis is accompanied by skewness till for the skew $D(s^2)$ III populations, the asymptotic value is attained even for small sample sizes. The asymptotic result is approached more slowly when kurtosis is large and for the very leptokurtic type IV and type VII distributions, sample sizes greater than twenty appear to be needed before the asymptotic values are closely approached.

7. M_1 AS A TEST FOR NORMALITY

The seriousness of the discrepancies to be expected, even in small samples, may be appreciated by considering M_1 as a *test* for normality.

Suppose a sample of N observations is drawn from a population distributed in the double χ form of equation (7). We shall assume that the mean η of the population is known, so that without loss of generality we may take it to be zero. We wish to test the hypothesis H_0 that the distribution is normal, i.e. that $\delta = 1$, $\sigma > 0$, against the class of alternatives H_1 which specify that the distribution is not normal, i.e. that $\delta \neq 1$, $\sigma > 0$. Then, following through the method set out by Neyman & Pearson (1933) for the selection of best critical regions, it is easy to show that the appropriate criterion is $M_{.1} = N \ln(\Sigma y^2/N) - \Sigma \ln y^2$, and that if $M_{.1}(\alpha)$ and $M_{.1}(1-\alpha)$ are the α and $1-\alpha$ significance points for $M_{.1}$, then the inequalities

$$M_{.1} > M_{.1}(\alpha) \quad \text{and} \quad M_{.1} < M_{.1}(1-\alpha) \tag{12}$$

define a pair of common best critical regions for testing H_0 against the alternative hypotheses $\delta < 1$ $(\gamma_2 > 0)$ and $\delta > 1$ $(\gamma_2 < 0)$, respectively, where in each case σ is unknown. Thus in the particular case when there is only one observation per group, $M_{.1}$ is itself the most sensitive criterion possible for testing for a certain type of non-normality. Non-normal distributions of the class considered, which are defined by equation (7), would rarely be expected to approximate to the distribution of actual data, but the above investigation suggests that we may form some idea of the sensitivity to non-normality of $M_{.1}$ for small samples by seeing how good it is as a *test* for non-normality.

---

* The actual standardized cumulants given above differ slightly in some cases from the theoretical values due to the finite size (10,000) of the populations which Le Roux sampled.

Table 3. *Characteristics of observed distributions of twenty-five values of M_1 for $k = 20$ groups of (a) $n = 20$*, (b) $n = 5$ observations, drawn from a variety of Pearson-type parent distributions*

			Mean			Variance			No. out of 25 exceeding normal theory significance point, P			
n	γ_1^2	γ_2	Expected, normal theory	Found in samples of 25 values of M_1	Expected for $D(s^2)$ III population with same γ_2	Expected, normal theory	Found in samples of 25 values of M_1	Expected $D(s^2)$ III population with same γ_2	$P = 10\%$	$P = 5\%$	$P = 1\%$	$P = 0{\cdot}1\%$
20	0·0	−0·5	19·4	15·0	14·7	39·4	16·5	22·7	0 (0%)	0 (0%)	0 (0%)	0 (0%)
20	0·5	0·7	19·4	25·8	25·9	39·4	92·7	70·8	10 (40%)	5 (20%)	3 (12%)	2 (8%)
20	1·0	0·8	19·4	28·6	26·9	39·4	93·3	76·1	13 (52%)	9 (36%)	4 (16%)	1 (4%)
25	1·2	2·8	19·4	37·4	46·5	39·4	159·0	227·2	21 (84%)	16 (64%)	10 (40%)	5 (20%)
20	0·0	4·1	19·4	39·4	59·1	39·4	329·4	367·1	19 (76%)	17 (68%)	11 (44%)	7 (28%)
5	0·0	−0·5	20·7	19·3	16·2	44·9	36·8	27·8	2 (8%)	1 (4%)	0 (0%)	0 (0%)
5	0·5	0·7	20·7	23·0	27·0	44·9	29·4	76·9	3 (12%)	2 (8%)	0 (0%)	0 (0%)
5	1·0	0·8	20·7	28·0	28·0	44·9	79·0	82·4	13 (52%)	9 (36%)	2 (8%)	0 (0%)
5	1·2	2·8	20·7	26·8	47·8	44·9	67·5	240·0	11 (44%)	8 (32%)	1 (4%)	0 (0%)
5	0·0	4·1	20·7	28·5	61·8	44·9	192·1	401·4	9 (36%)	5 (20%)	1 (4%)	1 (4%)

* In the single case $\gamma_1^2 = 1{\cdot}2$, $\gamma_2 = 2{\cdot}8$ the sample n number taken by Le Roux was 25 instead of 20.

(7·1) *Comparison of $M_{.1}$ with criteria for testing normality*

The usual test criterion for kurtosis is b_2, the sample estimate of the moment coefficient β_2. Geary (1935*a*) proposed as a new test for kurtosis the criterion

$$a = (\text{sample mean deviation})/(\text{sample standard deviation}).$$

This criterion had the advantage that unlike b_2, the distribution when the null-hypothesis of normality was true could be found fairly readily. Further investigation was carried out by Geary (1935*b*, 1947) to discover whether the power of this test to pick out a departure from normality was comparable to that of b_2. Pearson (1935) studied this question by means of sampling experiments. One such experiment consisted of drawing ten samples each of seventy-six observations from each of three symmetrical populations:

Description	γ_1^2	γ_2^*
Rectangular	0·0	−1·2
Pearson type VII	0·0	4·1
Double exponential	0·0	2·9

* As before the actual standardized cumulants given above differ slightly from the theoretical values for the last two populations due to the finite size (10,000) of the population sampled.

Pearson used his results to show the relative frequency with which the test criteria b_2 and a detected departures from normality and also to show the correlation between the results of the two tests. He published not only his conclusions but also the actual samples which he used and I have used these again to calculate the corresponding values for $M_{.1}$. In my calculation of $M_{.1}$ I have used the population mean (zero in each case), whereas in Pearson's calculations of b_2 and a the sample mean was used. In samples of seventy-six, however, the discrepancy this introduces should not be large. A further difficulty arises due to the fact that the samples are grouped to the nearest unit, resulting in a frequency class having zero deviation from the mean which would give rise to an infinite value of $M_{.1}$. To overcome this I have spaced the values evenly in the interval 0·0–0·5. Thus if there were n values in the zero frequency class these have been assumed to be at $1/(4n)$, $3/(4n)$, …, $(2n-1)/(4n)$. Fig. 2*a* shows the thirty values of $M_{.1}$ plotted against the corresponding values of b_2 and Fig. 2*b* shows the same values plotted against the values of a. It is seen that in each case marked correlation occurs, and if these graphs are compared with Pearson's Fig. 4 in which he plots a against b_2 it will be seen that the extent of correlation between $M_{.1}$ and b_2 and between $M_{.1}$ and a is similar to that between a and b_2. So far as is known, no tables are available for the significance points of M_1 (or for the equivalent criterion L_1) for the case of seventy-six groups with variance estimates each based on 1 degree of freedom. Also the asymptotic approximations of Bartlett (1937), Hartley (1940) and Box (1949) break down for large k and small numbers of degrees of freedom. Approximate significance points for $M_{.1}$ were therefore calculated by considering the distribution of $L_1^{-1} = \exp\{M_1/k\}$ which ranges from 1 to ∞. General expressions for the moments of L_1 were given by Neyman & Pearson (1931). From these the first four moments of L_1^{-1} were calculated and the appropriate Pearson type curve to approximate the distribution was found to be of type VI. It was assumed therefore that L_1^{-1} was approximately distributed as $cF_{\phi_1\phi_2}$, the curve starting at 1. By equating the first three moments of the two curves the values

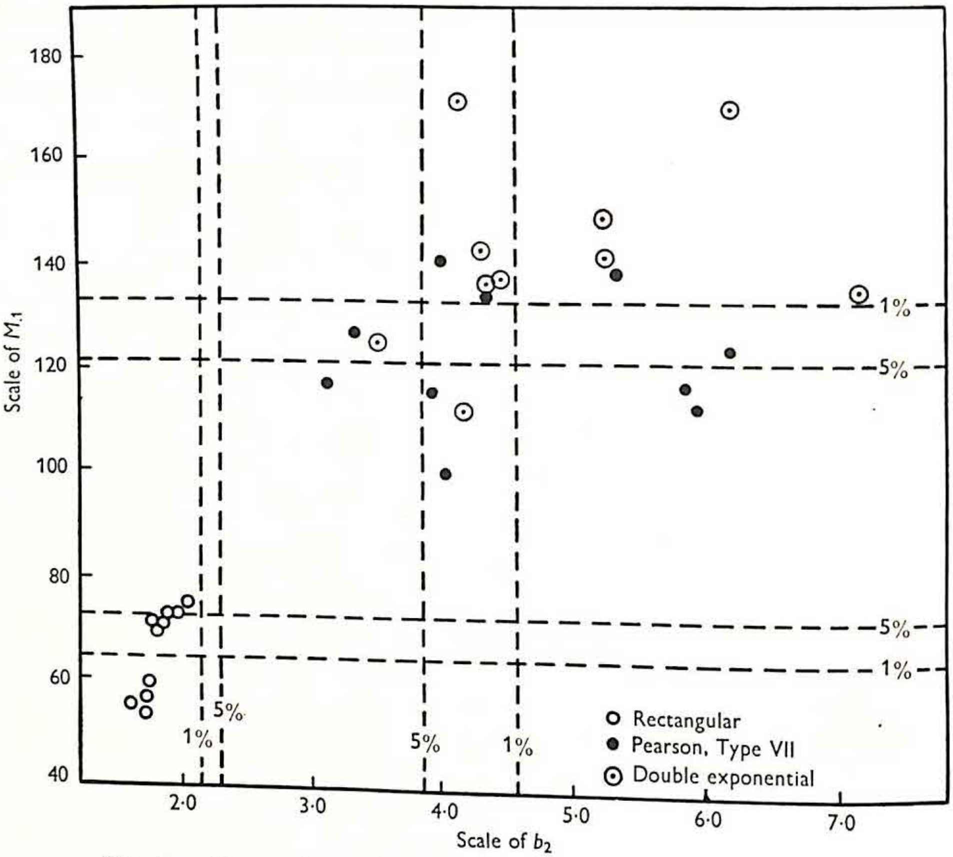


Fig. 2*a*. Comparison of $M_{.1}$ and b_2 in samples of seventy-six.

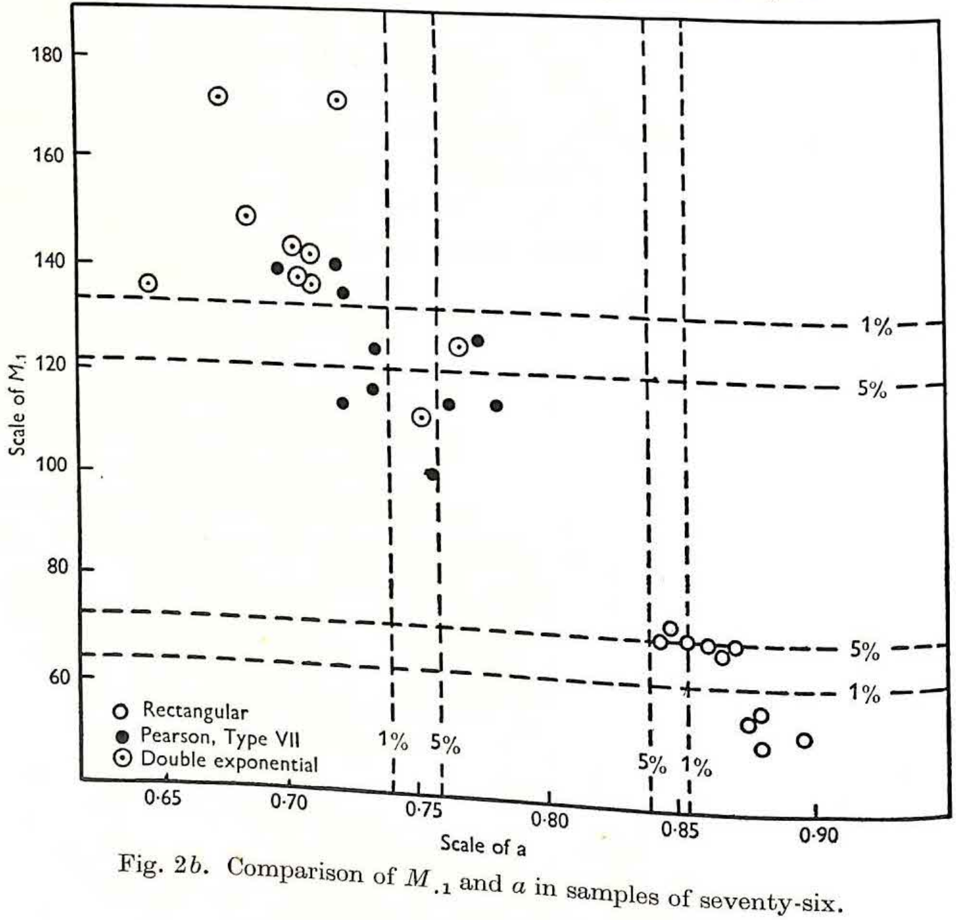


Fig. 2*b*. Comparison of $M_{.1}$ and a in samples of seventy-six.

$c = 2{\cdot}4705$, $\phi_1 = 67{\cdot}1998$, $\phi_2 = 45{\cdot}2904$ were obtained, and the approximate significance points calculated by two-way, four-point harmonic interpolation in the F tables of Merrington & Thompson (1943). The approximate significance points thus found were:

Lower		Upper	
1 %	5 %	5 %	1 %
64·1	72·3	121·2	133·3

Table 4 is an extension of Pearson's Table VII showing the number of values of b_2, a and $M_{.1}$ exceeding their 5 and 1 % significance levels.

Table 4. *Samples of seventy-six; comparison of criteria*

Distribution	Rectangular $\gamma_1^2 = 0{\cdot}0,\ \gamma_2 = -1{\cdot}2$			Pearson type VII $\gamma_1^2 = 0{\cdot}0,\ \gamma_2 = 4{\cdot}1$			Double exponential $\gamma_1^2 = 0{\cdot}0,\ \gamma_2 = 2{\cdot}9$		
	b_2	a	$M_{.1}$	b_2	a	$M_{.1}$	b_2	a	$M_{.1}$
No. of samples:									
within 5 % limits	0	0	3	2	3	5	1	1	1
between 5 and 1 % limits	0	2	3	4	1	2	5	1	1
beyond 1 % limits	10	8	4	4	6	3	4	8	8

For the rectangular and type VII distributions the criterion appears to be less sensitive to kurtosis than b_2 or a; for the double exponential distribution, however, $M_{.1}$ appears to be as sensitive as a and more sensitive than b_2. It is not of course contended that $M_{.1}$ is necessarily a practical test for kurtosis; we are concerned only to show that since b_2 and a are presumably very sensitive criteria for detecting kurtosis, $M_{.1}$ is much more sensitive *even when the sample size is small* than we should desire a test *for homogeneity of variances* to be.

In this connexion it is of some interest to reconsider some data discussed by Bartlett & Kendall (1946). The results appear as a two-way table with fifteen columns and three rows. The entries in the table are the logarithms of variance estimates s^2, each based on about 48 degrees of freedom. The authors show that by taking logarithms the data are brought to a suitable form for the application of analysis of variance. Their analysis is as follows:

	D.F.	S.S.	M.S.
Between rows	2	0·2667	0·1333
Between columns	14	0·1047	0·0075
Residual	28	0·1005	0·0036
Theoretical variance	—	—	0·0020

The first three items are calculated in the usual way, and the theoretical variance is based on the normal theory value $\kappa_2(\ln s^2) \sim 2/(\phi - 1)$. Compared with this theoretical variance the residual mean square is significant ($P = 0{\cdot}01$). It is deduced, therefore, that there is heterogeneity in the residual variance, and methods for further analysing this residual heterogeneity are discussed. Although familiarity with the distribution of their data (of which clearly a large quantity was available) no doubt rendered quite legitimate in this example the assumption by these authors of approximate normality, it is perhaps worth while emphasizing the dependence of this test on that assumption. For in general

$$\kappa_2(\ln s^2) \sim 2/(d\phi - 1),$$

consequently a value for γ_2 of 1·57 would have made the residual and theoretical variances exactly equal. Furthermore, since any value for the theoretical variance greater than 0·0024 would fail to show the residual variance significant at the 5 % level, any hypothesis that the residual variance is homogeneous and γ_2 greater than 0·44 is not contradicted by the data at this level of significance.

8. OTHER TESTS OF VARIANCE HOMOGENEITY

The reason for the difference in sensitivity to non-normality between the criterion M_2 for testing homogeneity of means and M_1 for testing homogeneity of variances can be seen as follows. In the analysis of variance (F-form) the criterion for means may be written as P/Q, where $P = \{\Sigma n_t(\bar{y}_t - \bar{y})^2\}/(k-1)$ is the between-groups mean square and $Q = s^2$ is the within-groups mean square. When the null hypothesis is true, P and Q each provides an unbiased estimate of σ^2 whether the observations are drawn from normal populations or not, so that Q is always a standard with which P may be usefully compared. But if in M_1, the criterion for variances, we write $s_t^2 = \sigma^2(1 + x_t)$ and formally expand the logarithms, we obtain $M_1 = \frac{1}{2}\sum_t \phi_t(x_t - \bar{x})^2$ plus terms in higher powers of x_t. For sufficiently large values of ϕ_t we can ignore the higher powers of x_t and we have $M_1/(k-1) \sim p/q$, where

$$p = \{\Sigma\phi_t(s_t^2 - s^2)^2\}/(k-1) \quad \text{and} \quad q = 2\sigma^4.$$

Comparing these with the expressions above we see that asymptotically the M_1 test is like an analysis of variance on the sample variances instead of sample means, *but the quantity p corresponding to the between-groups mean square is compared not with an estimate from the internal evidence of the samples but with a theoretical value of the variance which is appropriate only when the parent distribution is normal.* In general, the variance of a variance estimate s_t^2 is $d^{-1} \times 2\sigma^4/\phi_t$; consequently the appropriate value with which to compare p would be $d^{-1} \times 2\sigma^4$ and not $2\sigma^4$.

From the above discussion it will be seen that to obtain a criterion less dependent on γ_2, information on the variation to be expected in the sample variances gathered from internal evidence in the samples should be utilized. A test of variances should be 'studentized' for the fourth moment just as a test of means is studentized for the second moment.

(8·1) *Other tests not utilizing information within groups*

Other tests of variance homogeneity which do not utilize evidence on variance variability within the samples are equally sensitive to non-normality. In fact for $D(s^2)$ III parent distributions and, asymptotically, for all distributions with finite cumulants any criterion

$f\{s^2_{\phi_1}(\gamma_{21}), s^2_{\phi_2}(\gamma_{22}), \ldots, s^2_{\phi_k}(\gamma_{2k})\}$ will be distributed like $f\{s^2_{d_1\phi_1}(0), s^2_{d_2\phi_2}(0), \ldots, s^2_{d_k\phi_k}(0)\}$. Examples of such tests are those proposed by Cochran (1941) and by Hartley (1950). Cochran's criterion is $g = s^2_{\text{max.}}/\Sigma s^2_t$, whilst that proposed by Hartley is $F_{\text{max.}} = s^2_{\text{max.}}/s^2_{\text{min.}}$, where $s^2_{\text{max.}}$ and $s^2_{\text{min.}}$ are the largest and smallest of the group variances $s^2_1, s^2_2, \ldots, s^2_t, \ldots, s^2_k$. Using the same approximation as that adopted by Hartley it is a simple matter to calculate the true chance of exceeding the normal theory significance points for a $D(s^2)$ III parent population. For instance with k groups of 21 observations drawn at random from such a population in which $\gamma_2 = 1$ the approximate percentage chance of exceeding the normal theory 5 % significance point would be:

	$k = 5$	$k = 10$	$k = 20$
M_1	17·5	25·2	37·3
$F_{\text{max.}}$	17·3	23·3	30·3

The discrepancies using $F_{\text{max.}}$ are seen to be of similar magnitude to those found with M_1, and similar results are to be expected using Cochran's g. The multivariate tests for the constancy of the variance covariance matrix from one group of observations to another (Wilks, 1932; Bishop, 1939; Box, 1949, 1950) would be expected to be equally dependent on the assumption of multivariate normality.

Sensitivity to non-normality is found also with sequential tests. For example, Wald (1947) discussed the use of the sequential likelihood ratio test of the hypothesis H_0 that $\sigma^2 = \sigma^2_0$ when the alternative H_1 is that $\sigma^2 = \sigma^2_1$ and the observations $y_1, y_2, \ldots$ were drawn from normal populations with known mean η. Suppose the observations were drawn, not from a normal universe, but from the double χ population whose distribution is given by equation (7). Then it will be found that the logarithm of the likelihood ratio $L(\gamma_2)$ is equal to $\delta L(0)$ (where $L(0)$ is the logarithm of the likelihood ratio when the parent distribution is normal). Consequently if the quantity $L(0)$ is calculated when γ_2 is not zero and referred to limits $\ln(\beta/(1-\alpha))$ and $\ln((1-\beta)/\alpha)$, this is equivalent to referring the actual likelihood $\delta L(0)$ to the limits $\delta \ln\{\beta/(1-\alpha)\}$ and $\delta \ln\{(1-\beta)/\alpha\}$. The actual risks of error of the two kinds will therefore be α' and β' chosen so that $\beta'/(1-\alpha') = \{\beta/(1-\alpha)\}^\delta$ and $(1-\beta')/\alpha' = \{(1-\beta)/\alpha\}^\delta$. For $100\alpha = 100\beta = 5$ the percentage risk of errors of each of the two kinds is given below for various values of γ_2:

γ_2	-1	0	1	2
$100\alpha' = 100\beta'$	0·28	5·00	12·32	18·66

As with the tests with fixed sample sizes the result will be approximately true for any population with finite cumulants when the average sample sizes are not small. Roughly $\alpha' = \alpha^\delta$, $\beta' = \beta^\delta$ for small values of α and β. Similar difficulties would be expected with other sequential tests relating to the true value of and equality of variances; for example, with the tests proposed by Girshick (1946) and the tests based on ranges suggested by Cox (1949).

(8·2) *A test utilizing within-group information*

A test on variances less sensitive to parent non-normality must clearly utilize the within-group information in some way. Although the author believes that a better approach is available, one immediately practical method is to split up the groups of observations into sub-groups, and carry out an analysis of variance between and within groups on the logarithms of the sub-group variances, following Bartlett & Kendall (1946), who have shown the value of the logarithmic transformation in bringing variance data to a form suitable for the application of analysis of variance. The question of what sizes should best be taken for the subgroups requires further investigation.

Table 5. *Criteria calculated for ten samples of twenty groups of twenty observations drawn from a rectangular population*

Source of sample*		Test of means	Tests of variances			
				M_2 on logarithms of subgroup variances		
Page	Rows	M_2	M_1	2 sub-groups of 10	5 sub-groups of 4	10 sub-groups of 2
I	1–20	14·5	6·7	26·2	10·3	34·0
I	21–40	22·7	8·6	18·6	16·1	18·1
I and II	41–50 1–10	13·8	10·8	29·6	29·3	19·3
II	11–30	12·6	16·7	30·7	35·0	18·1
II	31–50	19·6	9·9	40·6	20·3	10·3
III	1–20	26·9	5·2	17·9	10·8	11·6
III	21–40	12·5	16·1	40·8	29·6	19·8
III and IV	41–50 1–10	18·0	7·4	34·1	21·6	34·1
IV	11–30	23·8	10·3	37·9	28·2	19·5
IV	31–50	11·8	10·3	30·6	32·4	29·6
Mean:	Found	17·6 ± 1·7	10·2 ± 1·2	30·7 ± 2·6	23·4 ± 2·8	21·4 ± 2·7
	Expected on normal theory	19·5	19·4	27·0	22·1	21·3
Variance:	Found	29·3 ± 15·6	13·9 ± 7·2	66·1 ± 35·3	78·7 ± 42·0	71·0 ± 37·9
	Expected on normal theory	40·1	39·4	79·0	51·5	48·1

* The groups of observations consisted in each case of the first twenty columns of numbers from pages of Fisher & Yates's tables. The page numbers and the numbers of the rows are shown above.

The results of a small sampling experiment of some interest in this connexion are set out in Table 5. Ten samples of twenty groups of twenty observations were drawn from the table of random numbers prepared by Fisher & Yates (1938). The parent distribution was thus effectively rectangular. As a test of group to group homogeneity of *means* a 'between and within groups' analysis of variance was performed for each of the ten samples of twenty groups and the ten resulting values of M_2 calculated. These are shown in the third column of the table. Four different tests for homogeneity of the variances were applied to each sample. The first was the M_1 test, the results for which are shown in the fourth column of the table. In the remaining three columns are shown the values of M_2 for analysis of variance performed on the logarithms of subgroup variances. The groups of twenty observations were divided into subgroups in three different ways: (i) two subgroups of ten observations, (ii) five subgroups of four, (iii) ten subgroups of two. The means and variances of the observed values together with their approximate standard errors and the values expected on normal theory are shown at the bottom of the table.

With so few samples, only large discrepancies of course would be detectable. We see that the test on means shows no evidence of departure from the values expected on normal theory even though the parent population is rectangular. The M_1 test for homogeneity of variance on the other hand shows extremely large departures, the mean being only about half the value expected on normal theory (for the rectangular parent distribution the asymptotic mean value is two-fifths of the normal-theory value). In contrast, it is seen that all the tests for homogeneity of variance based on M_2 give values agreeing fairly well with what would be expected if it could be assumed that the distribution of the logarithm of the variance was exactly normal, an assumption far from true, particularly for subgroups with only two observations.

9. DISCUSSION

It has frequently been suggested that a test of homogeneity of variances should be applied before making an analysis of variance test for homogeneity of means in which homogeneity of variance is assumed. The present research suggests than when, as is usual, little is known of the parent distribution, this practice may well lead to more wrong conclusions than if the preliminary test was omitted. It has been shown (Welch, 1937; David & Johnson, 1951*b*; Box, 1952; and Horsnell, 1953) that in the commonly occurring case in which the group sizes are equal, or not very different, the analysis of variance test is affected surprisingly little by variance inequalities. Since this test is also known to be very insensitive to non-normality it would be best to accept the fact that it can be used safely under most practical conditions. To make the preliminary test on variances is rather like putting to sea in a rowing boat to find out whether conditions are sufficiently calm for an ocean liner to leave port!

When the groups of observations were of unequal size and differences in variances might occur it would seem logical to replace the usual analysis of variance criterion which uses a pooled estimate of within-groups variance by the alternative criterion proposed by Welch (1951) and by James (1951), i.e. $\sum_t w_t(\bar{x}_t-\bar{x})^2$, where $w_t = n_t/s_t^2$. This criterion is robust to inequality of variance and almost certainly to non-normality also. (In fact, where inequality of variance might occur it would seem most logical to use Welch's criterion even if the groups were equal.)

When a criterion for testing a statistical hypcthesis is derived (for example, by the likelihood ratio method), it is usually necessary for purposes of mathematical convenience to over-simplify the specification of the problem. We should not be surprised therefore if an examination of the resulting criterion shows that the assumptions have sometimes, so to speak, been interpreted rather too literally. For this reason it is most important that derived criteria should be studied for robustness.

The property of robustness I believe to be even more important in practice than that the test should have maximum power and that the statistics employed should be fully efficient.

Where necessary I believe that the latter qualities should be sacrificed to ensure the former.* On the other hand, I do not think that we need necessarily go to the extreme of using non-parametric tests when it may well be that more powerful robust parametric tests can be found.

I am greatly indebted to Prof. E. S. Pearson for his interest and his many valuable suggestions for the improvement of the presentation of this paper. In conclusion, I wish to thank Mrs Margaret Edmondson for valuble assistance with the computations.

REFERENCES

BARTLETT, M. S. (1935). *Proc. Camb. Phil. Soc.* **31**, 223.
BARTLETT, M. S. (1936*a*). *Proc. Camb. Phil. Soc.* **32**, 560.
BARTLETT, M. S. (1936*b*). *Proc. Roy. Soc.* A, **154**, 124.
BARTLETT, M. S. (1937). *Proc. Roy. Soc.* A, **160**, 268.
BARTLETT, M. S. (1938). *Proc. Camb. Phil. Soc.* **34**, 33.
BARTLETT, M. S. & KENDALL, D. G. (1946). *J. R. Statist. Soc. Suppl.* **8**, 128.
BISHOP, D. J. (1939). *Biometrika*, **31**, 31.
BOX, G. E. P. (1949). *Biometrika*, **36**, 317.
BOX, G. E. P. (1950). *Biometrics*, **6**, 362.
BOX, G. E. P. (1952). Unpublished thesis, Ph.D., London University.
COCHRAN, W. G. (1941). *Ann. Eugen., Lond.*, **11**, 47.
COX, D. R. (1949). *J. R. Statist. Soc.* B, **11**, 101.
DAVID, F. N. & JOHNSON, N. L. (1951*a*). *Biometrika*, **38**, 43.
DAVID, F. N. & JOHNSON, N. L. (1951*b*). *Trabajos de estadistica*, **2**, 179.
FINCH, D. J. (1950). *Biometrika*, **37**, 187.
FISHER, R. A. & YATES, F. (1938). *Statistical Tables for Biological, Agricultural and Medical Research*, 1st ed. Edinburgh and London: Oliver and Boyd.
GAYEN, A. K. (1949). *Biometrika*, **36**, 353.
GAYEN, A. K. (1950*a*). *Biometrika*, **37**, 236.
GAYEN, A. K. (1950*b*). *Biometrika*, **37**, 399.
GEARY, R. C. (1935*a*). *Biometrika*, **27**, 310.
GEARY, R. C. (1935*b*). *Biometrika*, **27**, 353.
GEARY, R. C. (1947). *Biometrika*, **34**, 209.
GEARY, R. C. & PEARSON, E. S. (1938). *Tests of Normality*. Biometrika Office.
GIRSHICK, M. A. (1946). *Ann. Math. Statist.* **17**, 123.
HARTLEY, H. O. (1940). *Biometrika*, **31**, 249.
HARTLEY, H. O. (1950). *Biometrika*, **37**, 308.
HORSNELL, G. (1953). *Biometrika*, **40**, 128.
JAMES, G. S. (1951). *Biometrika*, **38**, 324.
KENDALL, M. G. (1943). *The Advanced Theory of Statistics*, **1**. London: Charles Griffin and Co.
LE ROUX, J. M. (1931). *Biometrika*, **23**, 134.
MERRINGTON, M. & THOMPSON, C. M. (1943). *Biometrika*, **33**, 73.
NEYMAN, J. & PEARSON, E. S. (1928). *Biometrika*, **20** A, 175 and 263.
NEYMAN, J. & PEARSON, E. S. (1930). *Bull. int. Acad. Cracovie*, A, p. 73.
NEYMAN, J. & PEARSON, E. S. (1931). *Bull. int. Acad. Cracovie*, A, p. 460.
NEYMAN, J. & PEARSON, E. S. (1933). *Phil. Trans.* A, **231**, 289.
PEARSON, E. S. (1931). *Biometrika*, **23**, 114.
PEARSON, E. S. (1935). *Biometrika*, **27**, 333.
TUKEY, J. W. (1948). *Human Biology*, **20**, 205.
WALD, A. (1947). *Sequential Analysis*. New York: John Wiley and Sons.
WELCH, B. L. (1937). *Biometrika*, **29**, 350.
WELCH, B. L. (1947). *Biometrika*, **34**, 28.
WELCH, B. L. (1951). *Biometrika*, **38**, 330.
WILKS, S. S. (1932). *Biometrika*, **24**, 471.

* Since writing the above a very interesting paper by J. W. Tukey (1948) has come to my notice which has many points of contact with the present paper and which expresses similar views to the above.

APPENDIX

Details of sampling experiments. Values of M_1 for $k = 20$ variance estimates, based on sets of $n = 20$, 25 and 5 observations randomly drawn from various parent populations

Parent population: Pearson type and values of (γ_1^2, γ_2)	II (0·0, −0·5)		III (0·5, 0·75)		I (1·0, 0·8)		IV (1·2, 2·8)		VII (0·0, 4·1)	
Sample	$n = 20$	$n = 5$	$n = 20$	$n = 5$	$n = 20$	$n = 5$	$n = 25$	$n = 5$	$n = 20$	$n = 5$
1	21·26	13·18	24·70	22·76	21·06	33·06*	50·74***	33·69*	94·48***	25·49
2	18·22	20·75	20·59	33·14*	42·46**	21·75	26·92	36·16*	44·32***	16·47
3	15·80	16·06	30·39?	35·76*	22·85	21·24	42·64**	33·95*	22·18	25·25
4	18·26	13·49	29·23?	18·00	20·70	13·28	80·00***	25·83	47·21***	32·19?
5	10·49	13·31	37·26**	19·34	33·93*	27·84	21·96	25·08	36·78*	14·58
6	12·66	22·95	23·84	18·29	57·62***	32·44?	18·70	32·27?	32·58*	17·41
7	10·87	19·16	16·64	28·49	42·16**	39·57**	35·44*	30·66?	40·81**	34·02*
8	19·83	32·19?	21·68	16·70	19·64	33·59*	42·77**	9·27	29·91?	36·64*
9	14·65	16·76	18·84	17·79	22·57	17·88	29·44?	19·40	24·42	28·46
10	13·63	14·90	19·42	30·27?	33·70*	38·44*	36·11*	15·59	36·09*	32·77*
11	18·43	20·12	16·94	25·75	30·41?	39·10*	35·14*	32·87*	37·86**	32·44?
12	20·27	17·61	27·26?	24·33	36·27*	33·83*	34·18*	18·44	15·45	24·58
13	10·84	14·88	28·72?	23·34	21·43	20·79	42·70**	33·08*	49·13***	25·01
14	15·98	27·92	18·77	17·01	37·74**	26·30	36·25*	28·50	37·89**	32·31?
15	21·22	20·55	31·54*	21·82	17·31	35·89*	52·25***	20·86	37·84**	21·01
16	12·53	14·30	19·39	22·08	31·55*	20·83	50·98***	27·67	24·68	17·64
17	8·35	17·80	30·62?	12·40	16·00	14·03	42·23**	35·38*	24·92	19·03
18	9·03	18·68	24·97	23·11	20·36	16·03	28·58?	42·13**	36·73*	27·42
19	16·55	17·26	19·97	29·34	20·30	30·89?	34·33*	26·62	72·17***	31·93?
20	10·07	18·12	13·09	25·70	21·16	46·49**	37·14**	28·14	57·89***	23·77
21	16·29	14·66	44·74***	24·12	28·55?	33·38*	45·22***	32·37?	31·22*	27·95
22	10·72	14·26	55·37***	25·27	31·44*	32·70?	30·32?	17·93	28·72?	87·95***
23	11·69	36·22*	32·86*	21·66	30·69?	30·22?	24·36	16·63	12·91	25·87
24	15·39	28·80	15·32	20·73	27·57	21·75	27·81?	15·16	71·40***	34·10*
25	21·61	21·05	22·30	19·05	28·70	19·75	29·25?	34·18*	36·56*	19·96

		10 %	5 %	1 %	0·1 %
Significance points for M_1 (Bartlett's approx.):	$n = 25$	27·60	30·58	36·72	44·46
	$n = 20$	27·70	30·70	36·86	44·63
	$n = 5$	29·58	32·78	39·36	47·65
Significance at these levels denoted by		?	*	**	***

APPROXIMATING TO THE DISTRIBUTIONS OF MEASURES OF DISPERSION BY A POWER OF χ^2

BY J. H. CADWELL, *Ordnance Board*

1. INTRODUCTION

The use of range in place of sample variance in various statistical tests has been the subject of a number of recent papers. In order to make range and means of sets of ranges amenable to certain lines of attack, two approximations have been proposed for the case of a normal parent population. The first consists in replacing range by a multiple of χ with an appropriate number of degrees of freedom. The two parameters at choice allow the mean and variance of range to be matched by those of the approximating form. This method has been discussed by Patnaik (1950), Hartley (1950) and Florin (1950); David (1951) gives tables of the necessary constants. The second approach, due to Cox (1949), replaces range by a χ^2 form. Pearson (1952) has compared the errors arising from the two methods. He points out that either can be regarded as an approximate transformation to the χ^2 form, and thus a number of standard tests are available. For instance, Bartlett's test for homogeneity of variance or the short-cut tables prepared by David (1952) may be used. Similarly, the ratio of a pair of ranges can be replaced by an approximate F-ratio.

Below we graduate the distribution of range by a multiple of a suitable power of χ^2. The extra parameter allows the first three moments to be matched, and much smaller errors result. By inverting the relationship an approximate transformation to the χ^2 form results. It should be noted that, while the χ approximation is dimensionally correct, the others are not. Thus the increased accuracy is obtained at the price of some loss of applicability of the approximation.

The method gives good results up to sample size 20 and the necessary constants are tabled. Results for the mean of a set of m ranges are also satisfactory and constants are given for m from 2 to 5 and n from 2 to 10. As there are no probability integral tables available for this statistic, these constants may be used to supply approximate percentage points from those of χ^2.

As the power of the statistic needed depends on sample size, the use of Bartlett's test or of an F-ratio test is restricted to values for samples of the same size. In this case David's 5 and 1 % short-cut tables can be used. These have been transformed into equivalent tables for the ratio of maximum range to minimum range. Enough values have been checked by quadratures to ensure that these latter tables are sufficiently accurate for most purposes.

In a paper to be published shortly, moment constants for the first quasi-range are given. This statistic is defined as the difference between the greatest-but-one and the least-but-one of the ordered values in the sample. For sample sizes beyond 17 it is more efficient than range as an estimator of standard deviation. It is also less influenced by the presence of an occasional 'rogue' observation, or by departures from the normal model. The same method of graduation gives good results and a table of constants is given.

Mean deviation can be graduated in the same way, and transformation constants and 5 and 1 % short-cut tables are provided.

Wilson & Hilferty (1931) have shown that, for large degrees of freedom, the cube root of χ^2 is almost normal. This leads to an approximate normalization of statistics that can be graduated by a power of χ^2. An example is given in the section on range.

Another transformation of some interest is the use of the logarithm of χ^2. This statistic has a variance independent of the parent value of the standard deviation, and can be used if it is desired to apply analysis of variance techniques to a set of observed standard deviations. However, while the statistic approaches normality in large samples, it does so rather slowly. Bartlett & Kendall (1946) suggest that the method should be used with caution for sample sizes below 10. We see that the logarithmic transformation when applied to range, or any of the other statistics considered below, will have the same properties. The variance of $\ln w$ will still be independent of the parent standard deviation and will be approximately

$$\alpha^2\left[\frac{d^2}{dx^2}\ln\Gamma(x)\right]_{x=\frac{1}{2}\nu}.$$

The errors introduced by the approximation to w will be small in comparison with the departures from normality of $\ln\chi^2$. Thus conditions for the use of the logarithmic transformation of range or mean deviation will be similar to those for its use with standard deviation.

It seems likely that other applications of this method of graduation may occur. Thus non-central χ^2 can be well approximated by a central χ^2 on a suitably chosen number of degrees of freedom. Still better results could be obtained by the present method. The use of frequency curves of the form

$$a+b(\chi^2)^\alpha \quad (\chi^2 \text{ on } \nu \text{ d.f.})$$

might also repay investigation.

2. DETERMINATION OF CONSTANTS

Consider the quantity

$$x = (\tfrac{1}{2}\chi^2)^\alpha,$$

where χ^2 is on ν degrees of freedom. The moments of x are readily found; they are given by the following relation:

$$\mu_r' = \frac{\Gamma(r\alpha+\frac{1}{2}\nu)}{\Gamma(\frac{1}{2}\nu)}. \tag{1}$$

Using Stirling's series, the expansions below follow at once:

$$V = \frac{\sqrt{\mu_2}}{\mu_1'} = \alpha\sqrt{\left(\frac{2}{\nu}\right)}\left\{1+\frac{(\alpha-1)^2}{2\nu}+\frac{(\alpha-1)^2(5\alpha-1)(\alpha-5)}{24\nu^2}+\dots\right\}, \tag{2}$$

$$\frac{\sqrt{\beta_1}}{V} = \frac{\mu_1'\mu_3}{\mu_2^2} = 3-\frac{1}{\alpha}+\frac{2(\alpha-1)^2}{\nu}+\dots. \tag{3}$$

On solving (2) and (3) for ν we obtain

$$\nu = \frac{2}{(3V-\sqrt{\beta_1})^2}\left\{1-\frac{(2V-\sqrt{\beta_1})^2(V+\sqrt{\beta_1})}{2(3V-\sqrt{\beta_1})}+\dots\right\}. \tag{4}$$

Inversion of (2) gives the series

$$\alpha = \alpha_0-\frac{\alpha_0(\alpha_0-1)^2}{2\nu}+\frac{\alpha_0(\alpha_0-1)^2(13\alpha_0^2+2\alpha_0+1)}{24\nu^2}+\dots, \tag{5}$$

where

$$\alpha_0 = \sqrt{\left(\frac{\nu}{2}\right)}\,V.$$

If the values of V and β_1 for a particular statistic u be inserted in (4) and (5), we shall have the results

$$u \text{ is approximately distributed like } (\chi^2/c)^\alpha, \tag{6}$$

$$cu^\lambda \text{ is approximately distributed like } \chi^2. \tag{7}$$

The constant c is chosen to make the means of the two variates in (6) agree. We have

$$\alpha = \frac{1}{\lambda} \quad \text{and} \quad \log c = \log 2 + \lambda\{\log\Gamma(\alpha+\tfrac{1}{2}\nu) - \log\Gamma(\tfrac{1}{2}\nu) - \log\bar{u}\}. \tag{8}$$

The expression for $\log c$ is readily evaluated using the log factorial tables prepared by Brownlee (1923).

For small ν or large α, (4) and (5) will not be sufficiently accurate. The following procedure was found to be satisfactory for range, quasi-range and mean deviation.

Since (5) comes directly from (2), the use of an approximate value for ν, while affecting the β_1 match of u and x does not affect the matching of mean and variance. Thus, for $\nu > 10$ it is found sufficient to round to the nearest integer and substitute this value in the first two terms of (5) to determine α.

When $10 > \nu > 4$ it is better to take ν to the nearest 0·1, and three terms of (5) are needed. In practice it was found preferable to solve

$$V^2+1 = \frac{\Gamma(2\alpha+\frac{1}{2}\nu)\,\Gamma(\frac{1}{2}\nu)}{\{\Gamma(\alpha+\frac{1}{2}\nu)\}^2}. \tag{9}$$

Thus, for mean deviation with $n = 5$, two terms of (5) give $\alpha = 0{\cdot}554$. Using (9) twice we find

$$\alpha = 0{\cdot}55, \quad V^2 = 0{\cdot}13655,$$
$$\alpha = 0{\cdot}56, \quad V^2 = 0{\cdot}14131.$$

Inverse linear interpolation for the required value of V^2 (0·13929) gives $\alpha = 0{\cdot}5558$. The use of trial values of α with a spacing of 0·01 obviates interpolation in Brownlee's tables and gives a maximum error in α of a few units in the fifth place.

When $\nu < 4$ the series (4) still gives very good results. Thus for range with $n = 3$ it gives 2·08, the true value being 2·05. However, it is now advisable to retain the second decimal in ν. Thus it becomes necessary to solve (9) and the relation

$$\sqrt{\beta_1}\,V^3 + 3V^2 + 1 = \frac{\Gamma(3\alpha+\frac{1}{2}\nu)\,\{\Gamma(\frac{1}{2}\nu)\}^2}{\{\Gamma(\alpha+\frac{1}{2}\nu)\}^3}. \tag{10}$$

This is readily done using trial values of ν rounded to 0·02 to obviate interpolation in the log factorial tables. Thus, for range with $n = 4$, (4) gives $\nu = 3{\cdot}23$. The values of V^2 and $\sqrt{\beta_1}$ at each of the four points

$$\nu = 3{\cdot}20, \quad \alpha = 0{\cdot}52, \qquad \nu = 3{\cdot}24, \quad \alpha = 0{\cdot}52,$$
$$\nu = 3{\cdot}20, \quad \alpha = 0{\cdot}53, \qquad \nu = 3{\cdot}24, \quad \alpha = 0{\cdot}53,$$

are then found using (9) and (10). Double inverse linear interpolation for the required values of V^2 and $\sqrt{\beta_1}$ then gives

$$\nu = 3{\cdot}202, \quad \alpha = 0{\cdot}5257.$$

With an interval of 0·04 in trial values of ν, neglect of second differences only affects α slightly in the fifth decimal place in the most unfavourable case.

It might be desirable to use a fixed value of λ for different sample sizes. This value could be chosen as a compromise between the values given for the various sample sizes in Table 1. The values of ν for each sample size will be found from

$$\nu = \nu_0 + (\alpha - 1)^2 - \frac{(\alpha^2 - 1)^2}{3\nu_0} + \dots, \tag{11}$$

where

$$\nu_0 = \frac{2\alpha^2}{V^2} \quad \text{and} \quad \alpha = \frac{1}{\lambda}.$$

When $\alpha = \frac{1}{2}$ this reduces to the series developed by Florin (1950) for the χ approximation. Two decimals in ν are usually adequate, and for $\nu_0 < 4$ we solve (9) for ν using trial values suggested by (11).

It sometimes happens that all but one or two samples are of the same size; the λ appropriate to this size may then be used throughout. For the exceptional samples ν and c are found from (11) and (8); for the standard size the values of Table 1 are used.*

3. RANGE

In order to apply an approximate test of homogeneity to a set of ranges, each for the same sample size, we first transform to approximate χ^2 values using the constants of Table 1. Then the usual Bartlett test is applied to these values. The adequacy of the approximate transformation to the χ^2 form was investigated as follows.

Tabular values of u, as near as possible to the percentage points listed below, were transformed by (7). The approximate probability integral was then found from tables of the χ^2 integral. This was compared with the exact value in tables of the probability integral of range prepared by Pearson & Hartley (1942). Errors in the approximate probability integral, expressed in units of the fourth decimal, are tabled below:

% point / n	0·5	5	20	50	80	95	99·5
3	−1	−2	+2	+3	−2	−2	−1
5	−1	−4	+5	+5	−3	−2	0
10	−3	−3	+5	+4	−4	−1	0
15	−2	0	+7	0	−3	0	+1
20	−2	+1	+5	−2	−3	+1	0

David's (1952) 5 and 1 % short-cut tables for the maximum F-ratio, transformed by (7), gave Tables 3 and 4. Exact values of ν, not the rounded values of Table 1, were taken. A four-point Lagrange formula was used when interpolating for ν, and harmonic values of ν were employed when it exceeded 10. For small ν the use of variance ratio became impossible, but the ratio of $s_{\text{min.}}$ to $s_{\text{max.}}$ was quite satisfactory for interpolation. The table for the 1 % values was not taken beyond $n = 10$ because interpolation and rounding errors begin to affect probabilities appreciably beyond this value.

We shall now apply a check to these values. If, in a set of k ranges, M denotes the ratio of maximum range to minimum range, we have

$$\text{Pr.}\,(M \geqslant M_0) = 1 - k\int_0^\infty \{F(M_0 x) - F(x)\}^{k-1} f(x)\,dx. \tag{12}$$

---

* All numbered Tables have been placed together at the end of the paper.

Using values of M_0 in Tables 3 and 4, the following probabilities were found by quadrature:

% point		$k = 2$	$k = 6$	$k = 11$
5 %	$n = 5$	0·0506	0·0514	0·0526
	$n = 10$	0·0508	0·0504	0·0516
	$n = 20$	0·0515	0·0516	0·0496
1 %	$n = 5$	0·0104	0·0106	0·0107
	$n = 10$	0·0103	0·0102	0·0094

These values show a slight bias, but are satisfactory for most purposes. In view of the size of the error at $n = 5$, $k = 11$, the true probability was also found at the nominal 5 % level for $n = 4$, $k = 11$; it was 0·0513. While the effect on probability of a change of $+0·01$ in the ratio varies over the tables, average values are $-0·0004$ for the 5 % tables and $-0·0002$ for the 1 % table.

We now turn to another application of the method. The Wilson-Hilferty approximation uses the fact that, for large ν,

$$\mathscr{E}\left(\frac{\chi^2}{\nu}\right)^{\frac{1}{3}} = 1-\frac{2}{9\nu}+O(\nu^{-3}), \quad \text{var.}\left(\frac{\chi^2}{\nu}\right)^{\frac{1}{3}} = \frac{2}{9\nu}+O(\nu^{-3}).$$

In addition, the distribution of this quantity is nearly normal. Applying this fact to range we find

$$w_5^{0·6099} \text{ is approximately Nr.}\{1·6443,\ 0·3844\},^*$$

$$w_{10}^{0·4881} \text{ is approximately Nr.}\{1·7163,\ 0·2197\}.$$

The errors in the approximate probability integral, found from the normal integral, are shown below in units of the fourth decimal:

% point	0·5	5	20	50	80	95	99·5
$n = 5$	$+20$	$+19$	-15	-16	-10	$+4$	-1
$n = 10$	$+1$	-1	-3	-1	$+2$	0	-1

While the approximation is good at $n = 10$, there are large errors when n is as small as 5.

4. MEAN RANGE

One method used when estimating dispersion in large samples consists in splitting the sample into subgroups. The mean of the subgroup ranges is then used as an estimator of standard deviation. Grubbs & Weaver (1947) have discussed the optimum method of splitting the sample. The mean of m ranges, each for the same sample size, will often arise naturally in other ways. Thus we might have m independent samples in which the variate has a fixed dispersion, but may take different mean levels in the various samples.

Table 2 gives the constants needed to transform the mean of m ranges, each for a sample of n, to the χ^2 form by means of (7). Except in one case, discussed later, no simple method is available for checking this approximation directly. Consequently the following procedure was used.

---

* Nr.$\{\mu, \sigma\}$ is used to denote a normal distribution, having mean $= \mu$ and standard deviation $= \sigma$.

While the two variates in (6) agree in their first three moments, the β_2 values will differ. Using (1) the following values were found for this β_2 difference. The β_2 of u was the larger of the two throughout:

n	$m = 1$	$m = 2$	$m = 3$	$m = 4$
2	0	−0·013	−0·012	−0·008
3	−0·005	−0·004	−0·002	−0·002
4	−0·010	−0·005	−0·003	−0·002
10	−0·015	−0·009	−0·006	−0·004

The percentage point tables prepared by Pearson & Merrington (1951) show that, for the region of the (β_1, β_2) plane concerned, a given β_2 error has a greater effect for large values of β_1 than for small; its effect varies but little with β_2. For the largest value of β_1 in the above table ($n = 2$, $m = 1$) the transformation is exact. The next largest, at $m = n = 2$, also shows the second largest β_2 difference.

The probability density function for a single range is of simple form when $n = 2$, and we find by integration that the mean of two such ranges has the density:

$$\frac{8}{\sqrt{(2\pi)}} e^{-\frac{1}{2}\bar{w}^2} \int_0^{\bar{w}} \frac{e^{-\frac{1}{2}x^2}}{\sqrt{(2\pi)}} dx.$$

The probability integral of $\bar{w}$ was found by quadrature, and the previous method of checking used to give the following errors in units of the fourth decimal:

% point	0·5	5	20	50	80	95	99·5
Error	−3	−5	+4	+11	−1	−4	+1

Since (β_1, β_2) approaches the normal point as m increases, this test, together with those made for $m = 1$ in the previous section, may be assumed to cover the least favourable cases.

5. THE FIRST QUASI-RANGE

In samples too large for the direct use of range, it may be necessary to randomize the observations before splitting them into groups. For instance, we may suspect that values are not independent of the order in which they are recorded. The lack of uniqueness may also be a drawback to the use of average range in some circumstances.

In such cases the first quasi-range, defined as the difference between the greatest-but-one and the least-but-one of the ordered values in the sample, may prove useful. Using the tables of expected values of range given by Tippett (1925) and reproduced in *Tables for Statisticians and Biometricians*. Vol. 2, the mean value of this statistic can be found from

$$\begin{aligned}\mathscr{E}(x_{n-1} - x_2) &= 2\mathscr{E}(x_{n-1}) = 2n(n-1)\int_{-\infty}^{\infty} F(1-F)^{n-2}\,dF \\ &= 2n(n-1)\int_{-\infty}^{\infty} (1-F)^{n-2}\,dF - 2n(n-1)\int_{-\infty}^{\infty} (1-F)^{n-1}\,dF \\ &= n\mathscr{E}(w_{n-1}) - (n-1)\,\mathscr{E}(w_n).\end{aligned}$$

Using the first three moments of quasi-range* the constants of Table 1 were found. These enable approximate percentage points to be deduced from those of χ^2 by using (6). Errors in the approximate probability integral of quasi-range found from (7) are given below in units of the fourth decimal:

% point / n	0·5	5	20	50	80	95	99·5
10	−1	−2	+1	+4	0	−1	0
20	−2	0	+6	+2	−4	−1	+1
30	−2	+1	+4	−2	−2	−1	+1

In order to avoid double interpolation in the Hartley-Pearson (1950) tables of χ^2, values of ν have been rounded to the nearest even integer when above 30.

6. MEAN DEVIATION

Work with this statistic was on the same lines as for range. Errors in the probability integral appear below; these are expressed in units of the fourth decimal:

% point / n	0·5	5	20	50	80	95	99·5
3	−1	−2	+2	+3	−2	−2	−1
4	−2	−5	+7	+10	−6	−3	+1
5	−2	−4	+8	+3	−6	−1	+1
6	−2	−3	+5	+3	−3	0	+1
10	−1	−1	+1	+3	−1	−1	0

As no tables are available beyond sample size 10, β_2 differences in (6) were again examined.

Series expansions for β_1 and β_2, due to Geary, are given by Pearson in an Editorial Note to a paper by Godwin & Hartley (1945). Godwin (1948) gives a minor correction to the β_2 series. Fisher (1920) derived the exact distribution of mean deviation when $n = 4$, and gave the values $\beta_1 = 0{\cdot}297$, $\beta_2 = 3{\cdot}28$. The series gives the values 0·299 and 3·252 respectively.

An examination of Fisher's paper shows that the fourth moment of σ_1 (the mean deviation estimator of σ) given there should be replaced by

$$\frac{9\pi}{512}(5\pi - 7\alpha) + \frac{165\pi}{384\sqrt{2}}, \quad (\alpha = \cos^{-1}\tfrac{1}{3}).$$

Using this value we obtain $\beta_1 = 0{\cdot}2983$, $\beta_2 = 3{\cdot}252$, in very good agreement with Geary's series.

The values of β_2 differences found are shown below:

n	3	4	5	6	10	15	20
β_2 difference	−0·005	−0·027	−0·017	−0·011	−0·008	−0·004	−0·003

---

* To be published shortly in a paper accepted for the *Annals of Mathematical Statistics*.

The large values at 4 and 5 are reflected in the errors found in the approximate probability integral. As mean deviation moves steadily towards normality with increase of n, we can expect errors beyond 10 to be small.

Tables 5 and 6 give the approximate upper 5 and 1 % values of the ratio of maximum mean deviation to minimum mean deviation in a set of k values. Using (12) the following values of probability were found from these tables by quadrature:

% point		$k = 2$	$k = 6$	$k = 11$
5 %	$n = 4$	0·0510	0·0520	0·0522
	$n = 10$	0·0506	0·0506	0·0509
1 %	$n = 4$	0·0104	0·0108	0·0111
	$n = 10$	0·0100	0·0101	0·0099

It will be seen that errors are similar to those for range. For sample sizes below 10 it will be found that the critical ratio differs very little for range, mean deviation or standard deviation.

The use of rounded values of ν in Table 1 obscures the fact that λ tends to a limit as n increases; this can readily be proved by using Geary's series. This suggests that, with a slight loss of accuracy, a fixed value of λ could be employed. It turns out that 1·8 is a satisfactory value, and appropriate tables are being prepared. The use of a fixed λ greatly increases the flexibility of the method by removing the restriction to a fixed sample size.

I should like to thank Mr D. F. Mills for his assistance in the preparation of the tables. I am indebted to Prof. Pearson for his advice on the presentation of this work, and for making an advance copy of the short-cut tables of critical variance ratios available.

Acknowledgement is made to the Chief Scientist, Ministry of Supply, for permission to publish this paper.

REFERENCES

BARTLETT, M. S. & KENDALL, D. G. (1946). *J.R. Statist. Soc. Suppl.* **8**, 128.
BROWNLEE, J. (1923). *Tracts for Computers*, **9**. Cambridge University Press.
COX, D. R. (1949). *J.R. Statist. Soc.* B, **11**, 101.
DAVID, H. A. (1951). *Biometrika*, **38**, 393.
DAVID, H. A. (1952). *Biometrika*, **39**, 422.
FISHER, R. A. (1920). *Mon. Not. R. Astr. Soc.* **80**, 768.
FLORIN, H. (1950). *Commun. Roy. Flem. Acad.* (Science Series), **12**, 6.
GODWIN, H. J. (1948). *Biometrika*, **35**, 304.
GODWIN, H. J. & HARTLEY, H. O. (1945). *Biometrika*, **33**, 252.
GRUBBS, F. E. & WEAVER, C. L. (1947). *J. Amer. Statist. Ass.* **42**, 224.
HARTLEY, H. O. (1950). *Biometrika*, **37**, 271.
HARTLEY, H. O. & PEARSON, E. S. (1950). *Biometrika*, **37**, 313.
HARTLEY, H. O. & PEARSON, E. S. (1951). *Biometrika*, **38**, 463.
PATNAIK, P. B. (1950). *Biometrika*, **37**, 78.
PEARSON, E. S. (1952). *Biometrika*, **39**, 130.
PEARSON, E. S. & HARTLEY, H. O. (1942). *Biometrika*, **32**, 302.
PEARSON, E. S. & MERRINGTON, M. (1951). *Biometrika*, **38**, 4.
TIPPETT, L. H. C. (1925). *Biometrika*, **17**, 364.
WILSON, E. B. & HILFERTY, M. M. (1931). *Proc. Nat. Acad. Sci., Wash.*, **17**, 694.

Table 1. *Transformation constants for range, mean deviation and first quasi-range*

n	Range			Mean deviation			n	First quasi-range		
	ν	λ	log c	ν	λ	log c		ν	λ	log c
							10	8·0	1·7289	0·3590
							11	9·3	1·7144	0·3884
2	1·00	2	$\bar{1}$·6990	1·00	2	0·3010	12	11	1·6683	0·4441
3	2·05	1·9619	$\bar{1}$·7632	2·05	1·9619	0·5766	13	12	1·6810	0·4477
4	3·20	1·9029	$\bar{1}$·8453	3·35	1·8437	0·7630	14	14	1·6271	0·5110
5	4·5	1·8298	$\bar{1}$·9393	4·6	1·7992	0·8851	15	16	1·5845	0·5648
6	6·0	1·7489	0·0412	5·9	1·7655	0·9823	16	18	1·5497	0·6117
7	7·7	1·6700	0·1426	7·2	1·7437	1·0612	17	20	1·5202	0·6532
8	9·5	1·6046	0·2315	8·4	1·7394	1·1235	18	22	1·4950	0·6908
9	12	1·5078	0·3543	9·6	1·7355	1·1778	19	24	1·4730	0·7250
10	14	1·4643	0·4215	11	1·7164	1·2326	20	27	1·4257	0·7857
11	16	1·4292	0·4794	12	1·7310	1·2695	21	29	1·4102	0·8129
12	19	1·3605	0·5748	13	1·7428	1·3036	22	32	1·3738	0·8629
13	22	1·3075	0·6537	15	1·6918	1·3596	23	36	1·3236	0·9290
14	26	1·2392	0·7521	16	1·7036	1·3874	24	38	1·3153	0·9482
15	30	1·1860	0·8338	17	1·7144	1·4136	25	42	1·2755	1·0036
16	34	1·1427	0·9036	18	1·7235	1·4383	26	46	1·2413	1·0530
17	40	1·0783	1·0030	20	1·6872	1·4799	27	50	1·2115	1·0977
18	46	1·0276	1·0858	21	1·6966	1·5012	28	54	1·1853	1·1384
19	52	0·9865	1·1569	22	1·7053	1·5215	29	58	1·1618	1·1758
20	60	0·9358	1·2431	23	1·7129	1·5409	30	62	1·1409	1·2104

If the statistic is u, then we have:

cu^{λ} has approximately the χ^2 distribution on ν degrees of freedom.

Table 2. *Transformation constants for the mean of m ranges, each for a sample of size n*

n	$m = 2$			$m = 3$		
	ν	λ	log c	ν	λ	log c
2	2·22	1·8376	0·1644	3·46	1·7822	0·3918
3	4·4	1·8539	0·1749	6·7	1·8282	0·3791
4	6·7	1·8339	0·2217	10	1·8302	0·4064
5	9·3	1·7822	0·2948	14	1·7740	0·4823
6	12	1·7385	0·3617	18	1·7337	0·5449
7	15	1·6849	0·4364	23	1·6633	0·6355
8	19	1·5995	0·5434	29	1·5838	0·7372
9	23	1·5375	0·6283	34	1·5477	0·7952
10	28	1·4620	0·7285	42	1·4613	0·9066

n	$m = 4$			$m = 5$		
	ν	λ	log c	ν	λ	log c
2	4·7	1·7562	0·5408	6·0	1·7310	0·6572
3	9·0	1·8152	0·5178	11	1·8339	0·6044
4	14	1·7816	0·5739	17	1·8054	0·6529
5	19	1·7547	0·6257	24	1·7446	0·7330
6	24	1·7319	0·6731	30	1·7310	0·7719
7	30	1·6812	0·7449	38	1·6683	0·8544
8	38	1·5959	0·8503	48	1·5870	0·9567
9	46	1·5354	0·9336	58	1·5288	1·0381
10	56	1·4605	1·0327	70	1·4607	1·1300

If the statistic is u, then we have:

cu^{λ} has approximately the χ^2 distribution on ν degrees of freedom.

Table 3. *Approximate upper 5 % points of the ratio of maximum value to minimum value in a set of k independent ranges. Each range is for a sample of size n from a normal population of fixed standard deviation*

k n	2	3	4	5	6	7	8	9	10	11	12
3	6·28	9·32	11·9	14·2	16·3	18·2	20·0	21·7	23·3	24·8	26·3
4	3·96	5·31	6·32	7·20	7·95	8·63	9·24	9·76	10·3	10·7	11·1
5	3·15	4·02	4·63	5·10	5·53	5·93	6·26	6·55	6·80	7·05	7·28
6	2·74	3·37	3·82	4·16	4·47	4·71	4·93	5·14	5·32	5·50	5·66
7	2·49	2·99	3·34	3·61	3·85	4·04	4·22	4·37	4·51	4·63	4·75
8	2·32	2·75	3·04	3·27	3·46	3·62	3·75	3·88	3·99	4·09	4·18
9	2·20	2·58	2·83	3·03	3·19	3·32	3·44	3·55	3·64	3·73	3·81
10	2·11	2·45	2·68	2·84	2·99	3·11	3·21	3·31	3·39	3·46	3·53
12	1·97	2·26	2·46	2·60	2·72	2·82	2·90	2·98	3·05	3·11	3·16
15	1·85	2·09	2·25	2·37	2·46	2·54	2·61	2·67	2·73	2·78	2·83
20	1·72	1·92	2·05	2·14	2·21	2·28	2·34	2·38	2·42	2·46	2·50

Table 4. *Approximate upper* 1 % *points for the range ratio*

$n \backslash k$	2	3	4	5	6	7	8	9	10	11	12
3	14·1	21·0	26·7	31·8	36·4	40·7	44·8	48·5	52·1	55·4	58·7
4	6·91	9·28	11·0	12·4	13·7	14·8	15·8	16·7	17·6	18·5	19·1
5	4·87	6·16	7·06	7·76	8·44	9·04	9·51	9·93	10·5	10·8	11·2
6	3·96	4·80	5·40	5·86	6·30	6·66	6·99	7·25	7·51	7·76	7·96
7	3·44	4·08	4·52	4·86	5·16	5·43	5·66	5·86	6·06	6·26	6·46
8	3·11	3·63	3·98	4·27	4·50	4·71	4·88	5·05	5·18	5·31	5·44
9	2·88	3·33	3·62	3·86	4·05	4·20	4·34	4·47	4·60	4·71	4·81
10	2·71	3·10	3·35	3·56	3·73	3·87	4·00	4·10	4·20	4·29	4·38

Table 5. *Approximate upper* 5 % *points for the ratio of maximum mean deviation to minimum mean deviation in a set of k independent values. Each value is derived from a sample of n from a normal population of fixed standard deviation*

$n \backslash k$	2	3	4	5	6	7	8	9	10	11	12
3	6·28	9·32	11·9	14·2	16·3	18·2	20·0	21·7	23·3	24·8	26·3
4	3·96	5·32	6·31	7·20	7·95	8·64	9·22	9·74	10·3	10·8	11·2
5	3·17	4·04	4·64	5·13	5·55	5·96	6·30	6·59	6·85	7·09	7·31
6	2·74	3·37	3·81	4·16	4·47	4·70	4·93	5·14	5·32	5·49	5·65
7	2·48	2·98	3·33	3·60	3·82	4·02	4·19	4·34	4·48	4·62	4·74
8	2·30	2·72	3·01	3·23	3·42	3·57	3·72	3·84	3·95	4·04	4·13
9	2·17	2·54	2·78	2·97	3·13	3·26	3·37	3·48	3·57	3·65	3·73
10	2·07	2·39	2·61	2·77	2·91	3·02	3·12	3·21	3·29	3·36	3·43
12	1·92	2·19	2·36	2·50	2·60	2·70	2·77	2·84	2·91	2·96	3·01
15	1·77	1·99	2·13	2·24	2·32	2·39	2·45	2·51	2·56	2·60	2·64
20	1·63	1·80	1·91	1·99	2·05	2·10	2·15	2·19	2·22	2·26	2·28
30	1·49	1·61	1·68	1·74	1·78	1·82	1·85	1·88	1·91	1·93	1·94
60	1·31	1·38	1·43	1·46	1·48	1·50	1·52	1·54	1·55	1·56	1·57
∞	1·00	1·00	1·00	1·00	1·00	1·00	1·00	1·00	1·00	1·00	1·00

Table 6. *Approximate upper* 1 % *points for the mean deviation ratio*

$n \backslash k$	2	3	4	5	6	7	8	9	10	11	12
3	14·1	21·0	26·7	31·8	36·4	40·7	44·8	48·5	52·1	55·4	58·7
4	6·92	9·30	11·0	12·3	13·6	14·8	15·9	16·8	17·6	18·4	19·0
5	4·89	6·18	7·12	7·83	8·49	9·04	9·50	10·0	10·5	10·9	11·2
6	3·96	4·79	5·41	5·86	6·30	6·66	6·98	7·26	7·52	7·75	7·96
7	3·43	4·07	4·51	4·85	5·15	5·41	5·64	5·85	6·06	6·26	6·44
8	3·08	3·60	3·95	4·22	4·45	4·67	4·84	5·00	5·15	5·28	5·40
9	2·83	3·27	3·56	3·80	3·98	4·15	4·29	4·42	4·53	4·63	4·72
10	2·65	3·02	3·27	3·47	3·64	3·77	3·89	3·99	4·09	4·18	4·26

THE POWER FUNCTION OF SOME TESTS BASED ON RANGE

By H. A. DAVID

1. Introduction

In a number of recent papers (Patnaik, 1950; Hartley, 1950*a*; David, 1951) tests based on the sample range have been developed to deal with the analysis of simple and double classifications as well as of some more complex orthogonal designs. It is our purpose to compare the power function of these short-cut tests with that of the analysis of variance procedure usually employed.

The new test criteria which may be used in place of the F-ratio are of two forms, depending on whether the test is one for a main factor or an interaction. In the former case the test function is the ratio of a single range to an independent mean range estimator of the population standard deviation σ, all ranges being taken over a *small* number of variates. In this situation a mean range can be closely approximated by a χ-distribution (with fractional degrees of freedom ν) so that the test ratio may be referred to tables of the 'Studentized' range q. The same χ-approximation allows us also to deal with the ratio of two independent mean ranges which arises in a test for an interaction.

The power of these tests has not been determined for any of the above designs, except in special cases. However, there has been evidence suggesting that the tests are likely to be reasonably powerful. Thus: (*a*) It is known that for a small sample size n the range provides a fairly efficient estimator of σ and that for large n ($n > 12$) a very good estimator of σ is given by the mean of ranges taken over small subgroups of the original sample (Davies & Pearson, 1934; Grubbs & Weaver, 1947). (*b*) For all of the above classifications tables are available giving the 'equivalent degrees of freedom' ν of estimates of error found from mean ranges (see, for example, David, 1951, pp. 408–9). ν is generally not much smaller than the number of error degrees of freedom in the analysis of variance. (*c*) Extensive investigations by quadrature methods made by Lord (1950) show the power of his u-test (Lord, 1947), employing a range estimate of variance, to be little less (on the systematic model) than that of the corresponding t-test, especially when mean ranges are employed.

These considerations suggest that the use of a *mean* range instead of a root mean square as a 'Studentizing' statistic leads to a definite but small loss in power. This will be confirmed in the sequel. It will also be shown that if a random probability model is relevant, range methods are somewhat less powerful than the standard procedure. For the systematic case attention is confined to a number of special alternate hypotheses. The q-test, whilst always retaining the property of unbiasedness (in the Neyman-Pearson sense) of the F-test, is found in certain rather special but important practical situations to be more powerful than the F-test.

2. The random model

We consider first the random set-up

$$x_{ij} = K + u_i + z_{ij} \quad (i = 1, \ldots, l;\ j = 1, \ldots, m), \tag{1}$$

where K is a constant and the u_i, z_{ij} are all mutually independent normal deviates with variances σ'^2 (for u) and σ^2 (for z). The presence of the u-terms in (1) makes the variance of the means $\bar{x}_i$ equal to $(\sigma^2+m\sigma'^2)/m$ instead of σ^2/m; hence the usual ratio of the between-groups mean square to the within-groups mean square is distributed not as F but as $(1+m\zeta^2)\,F$, where $\zeta = \sigma'/\sigma$, the degrees of freedom of F being $\nu_1 = l-1$, $\nu_2 = l(m-1)$. In this case the single-tailed F-test is known to be the uniformly most powerful test of the null hypothesis that $\zeta = 0$ against alternatives for which $\zeta > 0$.

When range methods are employed, we test the null hypothesis by referring the ratio

$$q_u = \frac{\sqrt{m}\,\text{range}\,(\bar{x}_i)}{\bar{w}/c} \tag{2}$$

to tables of percentage points of the 'Studentized' range q (Pearson & Hartley, 1943; May, 1952; Hartley, 1953). Since the presence of the u-terms multiplies the standard deviation of $\sqrt{m}\,\bar{x}_i$ by the factor $\sqrt{(1+m\zeta^2)}$, q_u will be distributed approximately as $\sqrt{(1+m\zeta^2)}\,q$, where q is the 'Studentized' range for sample size l and degrees of freedom ν.

Table 1. *Power of the q-test for a simple classification (random model); power of corresponding F-test* $= 1-\beta$

1*a*. $\alpha = 0{\cdot}05,\ 1-\beta = 0{\cdot}90$

$l \backslash m$	3	6	∞
4	0·90	0·90	0·90
6	0·86	0·88	0·89
8	0·85	0·87	0·87
10	0·83	0·86	0·86

1*b*. $\alpha = 0{\cdot}01,\ 1-\beta = 0{\cdot}90$

$l \backslash m$	3	6	∞
4	0·89	0·89	0·89
6	0·87	0·88	0·88
8	0·85	0·86	0·87
10	0·83	0·85	0·85

1*c*. $\alpha = 0{\cdot}05,\ 1-\beta = 0{\cdot}75$

$l \backslash m$	3	6	∞
4	0·74	0·74	0·74
6	0·68	0·72	0·73
8	0·67	0·70	0·71
10	0·65	0·69	0·69

1*d*. $\alpha = 0{\cdot}01,\ 1-\beta = 0{\cdot}75$

$l \backslash m$	3	6	∞
4	0·73	0·74	0·74
6	0·70	0·71	0·72
8	0·68	0·69	0·69
10	0·64	0·67	0·67

As pointed out by Patnaik (1950) the power function of the q-test can be obtained from tables of the probability integral of q and may be compared with the power of the corresponding F-test above. We do this by employing a general method for the comparison of the power function of two different tests of the same hypothesis (see Hartley, 1950*b*). For an assigned significance level α we find the value of ζ for which the power of the F-test has a selected value $1-\beta$; we then find the power of the q-test for the same value of ζ. If $F(\alpha;\nu_1,\nu_2)$ is the upper 100α % point of F for degrees of freedom ν_1, ν_2, it follows that the value of ζ is given by

$$1+m\zeta^2 = F(\alpha;\nu_1,\nu_2)/F(1-\beta;\nu_1,\nu_2).$$

Hence, using a similar notation for probability points of q, the power of the q-test for the same value of ζ is

$$\Pr.\,[\sqrt{(1+m\zeta^2)}\,q > q(\alpha;l,\nu)] = \Pr.\,[q > q(\alpha;l,\nu)\sqrt{\{F(1-\beta;\nu_1,\nu_2)/F(\alpha;\nu_1,\nu_2)\}}].$$

Results for the cases $\alpha = 0{\cdot}05$, $0{\cdot}01$; $\beta = 0{\cdot}10$, $0{\cdot}25$ and selected values of l and m are given in Table 1. Thus from Table 1 *a* we see, for example, that in testing for the significance of differences in means in eight categories, each containing six observations ($l = 8$, $m = 6$), a power of 0·90 given by the standard test is reduced to 0·87 when the present q-test is employed. It is evident from the table as a whole that the loss of power is small, especially in the commonest situations where the number of groups l is not very large nor the number within groups m very small. The effect of l on the power is clearly more marked than that of m. As $m \to \infty$ the problem reduces to that of comparing the range of the group means with the between-groups mean square as criteria for detecting departure of the between-groups population variance from $\sigma'^2 = 0$.

2·1. *A modified form of the test*

It is of some interest to determine the loss in power when only the denominator of the F-ratio is replaced by a range estimator (since presumably most of the reduction in power is due to the use of a single range in the numerator of q). The resulting test criterion is

$$m\Sigma(\bar{x}_i - \bar{x})^2/(\bar{w}/c)^2,$$

which is distributed approximately as F with degrees of freedom $l-1, \nu$. Using a significance level α, its power to detect a σ'^2 which the ordinary F-ratio test returns as significant with power $1-\beta$, is given approximately by the Incomplete Beta-ratio

$$\left.\begin{aligned} &I_t(\tfrac{1}{2}l(m-1), \tfrac{1}{2}(l-1)) = I_t(\tfrac{1}{2}\nu_1, \tfrac{1}{2}\nu_2), \\ \text{where}\qquad &t = \nu_2/(\nu_2+\nu_1 F'), \quad F' = F(\alpha;\, \nu_1, \nu)\, F(1-\beta;\, \nu_1, \nu_2)/F(\alpha;\, \nu_1, \nu_2). \end{aligned}\right\} \tag{3}$$

We have confined ourselves to the case $m = 3$ for which the drop in power due to the use of the 'Studentized' range criterion is most substantial. (To two-decimal accuracy the results are the same for $l = 6, 8, 10$.)

$l = 4$

α β	0·05	0·01
0·10	0·90	0·90
0·25	0·74	0·74

$l = 6, 8, 10$

α β	0·05	0·01
0·10	0·89	0·89
0·25	0·74	0·73

As expected, the loss in power is small, and there may be circumstances in which the present hybrid test criterion is of practical value.

2·2. *Test for a double classification*

Very similar results to those in Table 1 are obtained for the double classification

$$x_{ij} = K + u_i + v_j + z_{ij} \quad (i = 1, \ldots, l;\; j = 1, \ldots, m).$$

These are set out in Table 2, which is relevant to the power of the 'Studentized' range test for the existence of the u_i (the power of tests for the v_i follows by symmetry). m has been taken as 4, 7, ∞ so as to give F-ratios with the same degrees of freedom as corresponding entries in Table 1 with $m = 3, 6, \infty$.

2·3. *Power of modified F-ratio test for interaction*

In testing for an interaction a criterion of the form $F_1 = \left(\dfrac{\bar{w}_1/c_1}{\bar{w}_2/c_2}\right)^2$ arises through the use of range methods. F_1 is distributed approximately as an F-ratio with modified degrees of freedom ν_1', ν_2' which, as well as c_1, c_2, are obtainable from appropriate tables given by David (1951). It is clear that a power function comparison of the F_1- and F-tests may be made precisely as in (3). We have taken two cases typical of a split-plot experiment, but similar results will hold in general. We consider the comparison of 'error (1)' against 'error (2)', which is often carried out as a preliminary test of significance:

Table 2. *Power of the q-test for a double classification into l treatments and m blocks (random model)*

2*a*. $\alpha = 0{\cdot}05$, $1-\beta = 0{\cdot}90$

m l	4	7	∞
4	0·90	0·89	0·90
6	0·88	0·89	0·89
8	0·87	0·87	0·87

2*b*. $\alpha = 0{\cdot}01$, $1-\beta = 0{\cdot}90$

m l	4	7	∞
4	0·90	0·89	0·89
6	0·89	0·88	0·88
8	0·87	0·87	0·87

2*c*. $\alpha = 0{\cdot}05$, $1-\beta = 0{\cdot}75$

m l	4	7	∞
4	0·75	0·74	0·74
6	0·71	0·72	0·73
8	0·71	0·71	0·71

2*d*. $\alpha = 0{\cdot}01$, $1-\beta = 0{\cdot}75$

m l	4	7	∞
4	0·75	0·74	0·74
6	0·73	0·71	0·72
8	0·71	0·70	0·69

$l = 5,\ m = 4,\ n = 3$

α β	0·01	0·05
0·10	0·87	0·88
0·25	0·71	0·72

$l = 3,\ m = 3,\ n = 5$

α β	0·01	0·05
0·10	0·87	0·87
0·25	0·72	0·73

3. THE SYSTEMATIC SET-UP

We confine ourselves to the simple classification

$$x_{ij} = K + A_i + z_{ij} \quad (i = 1, \ldots, l;\ j = 1, \ldots, m),$$

where the systematic effects A_i are subject to the condition $\Sigma A_i = 0$. Similar results will be seen to hold for other models.

The case of two groups ($l = 2$) need not be considered here as the test ratio (2) is simply

$$\sqrt{m}\,|\,\bar{x}_1 - \bar{x}_2\,|/(\bar{w}/c),$$

and is equivalent to Lord's u-criterion appropriate to a two-sided test for the significance of the difference of two means (see § 1).

For $l > 2$ no simple procedure for the evaluation of the power of the q-test seems possible. We shall proceed by illustrating that, as in § 2·1, the loss in power due to the mean-range denominator alone is small, and by then investigating the more important changes in power introduced by the single-range numerator.

From Tang's (1938) tables it is easy to obtain the reduction in power when the degrees of freedom in the denominator of the F-ratio are reduced from $l(m-1)$ to ν, to give what we may term the F_R-test. We give two examples in the manner of Tang's table with

$$\phi = \sqrt{(m\Sigma A_i^2/l)}.$$

Example. (i) $l = 6$, $m = 6$; (ii) $l = 8$, $m = 3$.

Values of $1-\beta$

$\alpha = 0{\cdot}05$

ϕ	(i)		(ii)	
	F	F_R	F	F_R
1·5	0·738	0·728	0·711	0·696
2·0	0·952	0·948	0·942	0·937

$\alpha = 0{\cdot}01$

ϕ	(i)		(ii)	
	F	F_R	F	F_R
2·0	0·824	0·810	0·762	0·736
2·5	0·974	0·969	0·952	0·941

3·1. *The power of the range test*

A consideration of the numerator of q alone is clearly equivalent to an investigation of an ordinary range test, i.e. one for which σ is known. (2) now reduces to

$$\sqrt{m}\,\text{range}\,(\bar{x}_i)/\sigma,$$

a non-central range if the expectations of the $\bar{x}_i$ differ. We shall write, using a dash to denote non-centrality,

$$w' = \text{range}\,(y_i),$$

where the y_i follow normal distributions with expectations α_i (where $\Sigma\alpha_i = 0$) and unit variances.

Writing $z(x) = e^{-\frac{1}{2}x^2}/\sqrt{(2\pi)}$, the probability integral of w' is given by

$$P_l(w') = \sum_{i=1}^{l} \int_{-\infty}^{\infty} z(y_i-\alpha_i) \prod_{h\neq i}^{l} \left(\int_{y_i}^{y_i+w'} z(y_h-\alpha_h)\,dy_h \right) dy_i. \tag{4}$$

The power $(1-\beta)$ of the range test at significance level α, namely, $1-P_l(w_\alpha)$, is a symmetric function of the l expectations α_i.

In view of the constraint $\Sigma\alpha_i = 0$, the power is also a symmetric function of any $l-1$ of the α_i but does not appear to be expressible (unlike the corresponding probability integral of non-central χ^2 which depends only on $\Sigma\alpha_i^2$) as a function of less than these $l-1$ independent parameters (compare § 4).

We therefore confine attention to a few important special cases for which the evaluation of $P_l(w_\alpha)$ by approximate quadrature is practicable.

(a) *The case of a single outlying mean.* For this model we may write

$$\alpha_j = -\mu/l \quad (j = 1, \ldots, l-1), \qquad \alpha_l = (l-1)\,\mu/l.$$

From (4) we have after a simple translation

$$P_l(w_\alpha) = (l-1)\int_{-\infty}^{\infty} z(x) \left(\int_x^{x+w_\alpha} z(\xi)\,d\xi \right)^{l-2} \left(\int_x^{x+w_\alpha} z(\eta-\mu)\,d\eta \right) dx + \int_{-\infty}^{\infty} z(x-\mu) \left(\int_x^{x+w_\alpha} z(\xi)\,d\xi \right)^{l-1} dx.$$

The second term is very small, as it represents the contribution to $P_l(w_\alpha)$ when the outlier gives rise to the smallest observation. The power of the corresponding χ^2-test was obtained from the tables of non-central χ^2 given by Patnaik (1949) and Fix (1949) and from the $\nu = \infty$ line of Tang's tables, with $\lambda = \Sigma\alpha_i^2 = (l-1)\,\mu^2/l$ and $\phi = \sqrt{(\lambda/l)}$. In a few cases Patnaik's χ^2-approximation to χ'^2 had to be used.

Table 3. *Comparison of χ^2- and w-tests for a single outlying mean. Values of the power $(1-\beta)$*

α	ϕ \ l	2	6		10	
		χ^2, w	χ^2	w	χ^2	w
0·05	3	0·563	0·53	0·56	—	—
	4	0·807	0·82	0·83	0·76	0·80
	5	0·942	0·96	0·97	0·94	0·96
	6	0·989	—	—	0·994	0·996
0·01	4	0·600	0·62	0·63	0·54	0·61
	5	0·831	0·88	0·88	0·82	0·87
	6	0·952	0·98	0·98	0·97	0·98

It will be seen that except for $l = 2$, when the two tests are equivalent, the range test is slightly more powerful than the χ^2-test. This is not surprising in a test for outliers and a similar conclusion holds when, for example, there are two outliers, one at each end.*

(*b*) *The case of two equal clusters.* For simplicity take l even, then this set-up is defined by

$$\begin{aligned} \alpha_i &= -\tfrac{1}{2}\mu \quad (i = 1, \ldots, \tfrac{1}{2}l), \\ &= \tfrac{1}{2}\mu \quad (i = \tfrac{1}{2}l+1, \ldots, l), \end{aligned}$$

and $P_l(w_\alpha)$ given by

$$\begin{aligned} P_l(w_\alpha) = \tfrac{1}{2}l \int_{-\infty}^{\infty} z(x) \left(\int_x^{x+w_\alpha} z(\xi)\,d\xi\right)^{\frac{1}{2}l-1} \left(\int_x^{x+w_\alpha} z(\eta-\mu)\,d\eta\right)^{\frac{1}{2}l} dx \\ + \tfrac{1}{2}l \int_{-\infty}^{\infty} z(x-\mu) \left(\int_x^{x+w_\alpha} z(\xi)\,d\xi\right)^{\frac{1}{2}l} \left(\int_x^{x+w_\alpha} z(\eta-\mu)\,d\eta\right)^{\frac{1}{2}l-1} dx. \end{aligned}$$

A procedure similar to that in (*a*) yields the following comparisons:

$\alpha = 0{\cdot}05$

ϕ \ l	2	6		10	
	χ^2, w	χ^2	w	χ^2	w
3	0·563	0·823	0·74	0·94	0·82
4	0·807	0·981	0·95	0·999	0·98

For $l = 2$, this model is identical with (*a*). In other cases the range method is less powerful.

---

* Results reached by Dixon (1950) are of interest in this context. He plots for various tests for outliers, including the above w- and χ^2-tests, 'performance' curves obtained by sampling methods and finds the range test superior for all μ. However, his 'performance' is not measured in quite the same way as our 'power', as he discards all samples for which $\bar{x}_l$ is not the largest value in the sample.

4. SOME GENERAL COMMENTS

We now consider to what extent the conclusions regarding the range test of § 3·1 can be converted into conclusions about 'Studentized' range tests. Case (*a*) is of interest in suggesting that when the models there discussed are relevant a better test than the standard one of the analysis of variance can be obtained by replacing the F-ratio by

$$\sqrt{m}\,\text{range}\,(\bar{x}_i)/s,$$

where $$s = \sqrt{\{\Sigma\Sigma(x_{ij}-\bar{x}_i)^2/l(m-1)\}},$$

giving a 'Studentized' range test with root-mean-square denominator. The test criterion range $(\bar{x}_i)/(\bar{w}/c)$ may also lead to a more powerful test, especially when the number of degrees of freedom in the numerator is large, but the issue is somewhat obscured by the small loss in power due to the use of $\bar{w}/c$ in place of s. This does not, of course, contradict certain optimum properties of the analysis of variance tests proved by Hsu (1941) and Wald (1942). However, it shows that the non-central range distribution does not depend on $\Sigma\alpha_i^2$ alone, as otherwise Hsu's result would not permit it to lead to a more powerful test than the F-ratio.

These comments provide no rigorous conclusions. It would be possible, following lines similar to those of Hartley (1948), to obtain exact upper bounds for the difference between the power functions of the F- and q-tests in terms of the maximum difference between the power functions of the χ and range tests. But as the distribution of non-central range is not available, this approach cannot at present lead us to any numerical gauge of accuracy.

My thanks are due to Dr H. O. Hartley for his guidance in this work which was carried out at University College, London.

REFERENCES

DAVID, H. A. (1951). *Biometrika*, **38**, 393.
DAVIES, O. L. & PEARSON, E. S. (1934). *J.R. Statist. Soc. Suppl.* **1**, 76.
DIXON, W. J. (1950). *Ann. Math. Statist.* **21**, 488.
FIX, EVELYN (1949). *Univ. Calif. Publ. Statist.* **1**, 15.
GRUBBS, F. E. & WEAVER, L. C. (1947). *J. Amer. Statist. Ass.* **42**, 224.
HARTLEY, H. O. (1948). *Biometrika*, **35**, 417.
HARTLEY, H. O. (1950*a*). *Biometrika*, **37**, 271.
HARTLEY, H. O. (1950*b*). *Biometrika*, **37**, 308.
HARTLEY, H. O. (1953). *Biometrika*, **40**, 236.
HSU, P. L. (1941). *Biometrika*, **32**, 62.
LORD, E. (1947). *Biometrika*, **34**, 41.
LORD, E. (1950). *Biometrika*, **37**, 64.
MAY, JOYCE M. (1952). *Biometrika*, **39**, 192.
PATNAIK, P. B. (1949). *Biometrika*, **36**, 202.
PATNAIK, P. B. (1950). *Biometrika*, **37**, 78.
PEARSON, E. S. & HARTLEY, H. O. (1943). *Biometrika*, **33**, 89.
TANG, P. C. (1938). *Statist. Res. Mem.* **2**, 126.
WALD, A. (1942). *Ann. Math. Statist.* **13**, 434.

SOME SIMPLE APPROXIMATE TESTS FOR POISSON VARIATES

By D. R. COX

Statistical Laboratory, University of Cambridge

1. Introduction

If events, such as accidents or stoppages of a machine, occur randomly in time at a true rate λ, the number of events in a fixed time t follows a Poisson distribution of mean λt. In inverse sampling, with n fixed, the time t up to the nth event is distributed as $(2\lambda)^{-1}\chi^2_{2n}$, where χ^2_{2n} denotes a χ^2 variate with $2n$ degrees of freedom. Barnard (1946) has pointed out that the latter result leads to convenient 'sequential' tests of hypotheses about λ. For example, if we have inverse samples (n_1, t_1) and (n_2, t_2) from two populations, the hypothesis that $\lambda_1 = \lambda_2$ is tested by referring $F = t_1 n_2/t_2 n_1$ to the variance-ratio tables with $(2n_1, 2n_2)$ degrees of freedom. In the present note we show that, with a slight modification, Barnard's tests apply almost exactly to direct Poisson sampling in which t is fixed and n is a random variable. We obtain, in particular, a convenient test for the equality of Poisson means, although the main use of the method is likely to be in more complicated situations where λ is the product of unknown parameters.

2. The basic approximation

In direct Poisson sampling in which the number of events occurring in a fixed time is recorded, we have

$$\text{prob (no. of events} \geqslant n) = \sum_{r=n}^{\infty} \frac{e^{-\lambda t}(\lambda t)^r}{r!} = \text{prob}\left(\frac{1}{2\lambda}\chi^2_{2n} < t\right), \tag{1}$$

and

$$\text{prob (no. of events} \geqslant n+1) = \text{prob}\left(\frac{1}{2\lambda}\chi^2_{2n+2} < t\right). \tag{2}$$

If we wish to make an approximation to prob (no. of events $> n$) in which the number of events is treated as a continuous variate, it is reasonable to take a quantity intermediate between (1) and (2). A natural choice is

$$\text{prob (no. of events} > n) \simeq \text{prob}\left(\frac{1}{2\lambda}\chi^2_{2n+1} < t\right),$$

i.e. we calculate probabilities as if

$$2\lambda t \text{ is distributed as } \chi^2_{2n+1}. \tag{3}$$

Thus the suggestion is that if we observe n events in a fixed time t, we may work approximately as if n is fixed and $2\lambda t$ is distributed as χ^2_{2n+1}. The remainder of the paper is concerned with the consequences of (3).

3. Comparison of two populations

If we sample two populations with rates of occurrence λ_1, λ_2 and in times t_1, t_2 observe n_1, n_2 events, then, according to (3),

$$\frac{t_1(n_2+\frac{1}{2})}{t_2(n_1+\frac{1}{2})}\frac{\lambda_1}{\lambda_2} \tag{4}$$

is distributed approximately as F with $(2n_1+1, 2n_2+1)$ degrees of freedom. Thus we may test the hypothesis that $\lambda_1 = \lambda_2$ by referring

$$F = \frac{t_1(n_2+\frac{1}{2})}{t_2(n_1+\frac{1}{2})} \tag{5}$$

to the F tables with $(2n_1+1, 2n_2+1)$ degrees of freedom. Also $(100-2\alpha)\,\%$ confidence intervals for λ_1/λ_2 may be obtained from

$$\frac{t_2(n_1+\frac{1}{2})}{t_1(n_2+\frac{1}{2})}F_- < \frac{\lambda_1}{\lambda_2} < \frac{t_2(n_1+\frac{1}{2})}{t_1(n_2+\frac{1}{2})}F_+, \tag{6}$$

where F_-, F_+ are the lower and upper $\alpha\,\%$ points of F with $(2n_1+1, 2n_2+1)$ degrees of freedom.

Example. Observations of the spinning of one batch of wool gave 5 ends down in 200 spindle hours and of a second batch 12 ends down in 180 spindle hours.

Assuming that the occurrence of ends down is random for each batch, we may test for the difference between batches by

$$F = \frac{200}{180} \times \frac{12\cdot5}{5\cdot5} = 2\cdot53,$$

with (11, 25) degrees of freedom. The 5 % point in the ordinary F tables is 2·20, the $2\frac{1}{2}$ % point 2·56. Thus in a two-sided test the difference between batches is very nearly significant at 5 %. The lower and upper $2\frac{1}{2}$ % points of F with (11, 25) degrees of freedom are 1/3·16 and 2·56, so that by (6) a 95 % confidence interval for the ratio of the stoppage rates is

$$\frac{180 \times 5\cdot5}{200 \times 12\cdot5} \times \frac{1}{3\cdot16} < \frac{\lambda_1}{\lambda_2} < \frac{180 \times 5\cdot5}{200 \times 12\cdot5} \times 2\cdot56,$$

i.e.

$$0\cdot125 < \lambda_1/\lambda_2 < 1\cdot01.$$

The test of the difference between batches could equally well be done by the conventional χ^2 method, but the provision of a simple confidence interval for the ratio of the stoppage rates is a useful additional feature of the present approach.*

4. ACCURACY OF THE TEST

The test (5) may be expected to be accurate for large samples; its accuracy for small samples may be investigated as follows. The distribution of n_1, n_2 corresponding to given λ_1, λ_2, t_1 and t_2 is exactly

$$\text{prob}\,(n_1, n_2) = e^{-\lambda_1 t_1}\frac{(\lambda_1 t_1)^{n_1}}{n_1!}\,e^{-\lambda_2 t_2}\frac{(\lambda_2 t_2)^{n_2}}{n_2!}. \tag{7}$$

If we first determine the critical region of the test (5), we can then find the exact probability associated with the test by adding (7) over all points in the critical region. Table 1 gives the results of such calculations. In all cases a one-sided test of the hypothesis $\lambda_2 = \lambda_1$ against alternatives $\lambda_2 > \lambda_1$ has been examined.

The general conclusion from Table 1 is that except when the population means are very small, the approximate F test gives the probability of errors of the first kind sufficiently accurately for practical purposes. For samples of the same size, the test may be considered satisfactory at the 0·05 level if the true mean exceeds one, and satisfactory at the 0·01 level if the true mean exceeds two.

---

* *Note added in proof.* The derivation of confidence intervals for the ratio of Poisson means has recently been considered in detail by Chapman (1952).

It is natural to compare the approximate F test with the exact test of Przyborowski & Wilenski (1940), which is based on the distribution of n_1, n_2 conditional on n_1+n_2. The comparison is difficult because the latter test is discrete and the true size of the critical region at, say, the 5 % level is appreciably less than 5 %, when small numbers are involved. The critical region of the approximate F test appears always to consist of the critical region of the exact test with some additional points. It is thus roughly equivalent to Barnard's C.S.M. test (Barnard, 1947).

Table 1. *Exact probabilities associated with various nominal significance levels of the approximate F test of the hypothesis $\lambda_2 = \lambda_1$ against alternatives $\lambda_2 > \lambda_1$*

(*a*) *Samples of equal size, $t_1 = t_2$*

Population mean, $\lambda_1 t_1 = \lambda_2 t_2$	Nominal significance level			
	0·1	0·05	0·01	0·001
1	—	0·031	0·001	—
1·5	—	0·049	0·004	—
2	0·124	0·059	0·007	0·0001
3	—	0·065	0·011	—
4	—	0·062	0·012	—
5	—	0·058	0·012	—
6	0·101	0·054	0·012	0·0013

(*b*) *First sample size the larger, $t_1 = 3t_2$*

Smaller population mean, λt_2	Nominal significance level	
	0·05	0·01
1	0·048	0·008
1·5	0·055	0·009
2	0·055	0·010
3	0·056	0·011
4	0·056	0·010
5	0·051	0·010

(*c*) *Second sample size the larger, $t_2 = 3t_1$*

Smaller population mean, λt_1	Nominal significance level	
	0·05	0·01
1	0·012	3×10^{-5}
1·5	0·038	5×10^{-4}
2	0·056	0·002
3	0·059	0·010
4	0·059	0·013
5	0·057	0·012

The approximate F test is an interesting consequence of (3) and may sometimes be preferred to conventional approximate χ^2 methods. The detailed discussion does, moreover, show that (3) may lead to accurate results even in very small samples; in § 7 we shall consider an application where a great simplification is achieved by the use of (3).

5. Confidence intervals for a single mean

The approximation (3) may be used to test the hypothesis that $\lambda = \lambda_0$, or to obtain confidence intervals for λ. An interval of confidence coefficient $(100-2\alpha)$ % is thus given approximately by

$$\tfrac{1}{2}\chi^2_{2n+1,-} < \lambda t < \tfrac{1}{2}\chi^2_{2n+1,+}, \tag{8}$$

where $\chi^2_{\nu,\pm}$ are the upper and lower α % points of χ^2 with ν degrees of freedom. This may be compared with Garwood's (1936) confidence interval, which in the present notation is

$$\tfrac{1}{2}\chi^2_{2n,-} < \lambda t < \tfrac{1}{2}\chi^2_{2n+2,+}. \tag{9}$$

The interval (9) has a confidence coefficient of *at least* $(100-2\alpha)$ % whatever the true mean λt; the true confidence coefficient is a serrated function of λt which, for small λt, has most of its values appreciably greater than $(100-2\alpha)$ %.

The interval (8) is always narrower than (9), and the true confidence coefficient is sometimes less than $(100-2\alpha)$ %. In many practical applications it would be reasonable to assume that over a number of applications of the method the true means λt are distributed randomly with respect to the serrations of the graph of the confidence coefficient. In this case it would be justifiable to use a confidence interval for which the average confidence coefficient over any fairly small range of λt is at least $(100-2\alpha)$ %. The interval (8) appears to satisfy this condition provided that n and α are not very small, and it might therefore be claimed that (8) is preferable to (9). The point is, however, of little practical importance because the reduction in width from (9) to (8) is only small.

Example. Seven stoppages of a machine are observed in a certain period. 90 % confidence intervals for the population number of stoppages are, according to (8), (3·63, 12·50) and according to (9), (3·29, 13·15).

6. Relation with inverse sampling

The close connexion between the tests given in §§ 3 and 5 and those based on inverse sampling has been mentioned briefly in §§ 1 and 2. This connexion will now be discussed in more detail.*

Tests based on the measurement of the intervals between randomly occurring events have been described in full by Maguire, Pearson & Wynn (1952). In particular, their test for the significance of the difference between the rates of occurrence in two series is as follows. Let n_1, n_2 be fixed and let t_1, t_2 be random variables defined as the times in the two series up to the n_1, n_2th events. Then $F = t_1 n_2 / t_2 n_1$ may be tested *exactly* as a variance ratio with $(2n_1, 2n_2)$ degrees of freedom. The test of § 3, on the other hand, is that if t_1, t_2 are fixed times and if random variables n_1, n_2 are defined as the numbers of events occurring in times t_1, t_2, then $F = t_1(n_2+\frac{1}{2})/t_2(n_1+\frac{1}{2})$ may be tested *approximately* as a variance ratio with $(2n_1+1, 2n_2+1)$ degrees of freedom.

---

* I am very grateful to Prof. E. S. Pearson for some helpful comments leading to the addition of this section.

In the first case the interval t_1 ends with the n_1th event, but in the second case this is not in general so. The extra degrees of freedom, and the $\frac{1}{2}$'s in the definition of F, may be thought of as accounting for the periods between the last events and the close of the periods of observation. In many applications in which t_1, t_2 are fixed, the precise instants at which the n_1, n_2th events occur would not be known.

It will frequently happen in applications that neither t_1, t_2 nor n_1, n_2 are fixed in advance of the observations. If, however, the periods t_1, t_2 are determined by some random process that is quite uninfluenced by the numbers of events occurring, then conditionally on t_1, t_2, the quantities n_1, n_2 follow Poisson distributions and the approximate F test may be applied. Similarly, if the time intervals are measured up to the n_1, n_2th events, where n_1, n_2 are determined by a random process independent of the observations, then the exact F test with $(2n_1, 2n_2)$ degrees of freedom is applicable.

There are many other possible ways in which t_1, t_2 might be determined; for example, t_1, t_2 might be chosen by some random process correlated with, but not completely determined by n_1, n_2. In such cases it is not possible to find the properties of the above tests without special investigation, although a reasonable general rule is to use the first F test, with $(2n_1, 2n_2)$ degrees of freedom, whenever the intervals t_1, t_2 are ended by events, and the second test otherwise. This rule breaks down in extreme cases; for example, if t_1, t_2 are determined by a likelihood-ratio sequential test for comparing the rates of occurrence, it would clearly be entirely inappropriate to apply either F test.

7. THE LOGARITHMIC TRANSFORMATION

The basic approximation (3) is very conveniently expressed in logarithmic form, and so may be expected to be useful whenever λ is the product of unknown parameters. If f_n is any function of n we may rewrite (3)

$$\log\frac{f_n}{t} = \log\lambda + \log f_n - \log\tfrac{1}{2}\chi^2_{2n+1}.$$

The $\log\chi^2$ distribution has been studied in detail by Bartlett & Kendall (1946) and by Wishart (1947), who have, in particular, shown that

$$E\log\tfrac{1}{2}\chi^2_\nu = \psi(\tfrac{1}{2}\nu), \quad \operatorname{var}\log\tfrac{1}{2}\chi^2_\nu = \psi'(\tfrac{1}{2}\nu),$$

where ψ, ψ' are the digamma and trigamma functions. It is convenient to choose f_n so that $\log f_n/t$ is an unbiased estimate of $\log\lambda$. Thus we take $f_n = \exp\{\psi(n+\frac{1}{2})\}$, and for the present purpose it is accurate enough to write

$$f_0 = 0{\cdot}14, \quad f_n = n \quad (n = 1, 2, \ldots). \tag{10}$$

Also

$$\left.\operatorname{var}\log\frac{f_n}{t} = \psi'(n+\tfrac{1}{2}) \simeq \begin{array}{ll} 4{\cdot}93 & (n=0), \\ 1/n & (n = 1, 2, \ldots). \end{array}\right\} \tag{11}$$

Thus if we define a transformed variate by

$$\left.\begin{aligned} z &= \log_{10}\frac{0{\cdot}14}{t} \quad \text{if} \quad n = 0, \\ &= \log_{10}\frac{n}{t} \quad\quad \text{if} \quad n \neq 0, \end{aligned}\right\} \tag{12}$$

then approximately $E(z) = \log_{10}\lambda$ and $\operatorname{var} z = (\log_{10} e)^2 v_i$,

where

$$\left.\begin{aligned} v_i &= 4{\cdot}93 \quad (n=0),\\ &= 1/n \quad (n \neq 0).\end{aligned}\right\} \tag{13}$$

One use of the transformation will now be described.

Suppose observations of stoppage rates are made for k machines on each of which stoppages occur at random, the rates for the different machines being different. Suppose that a change in the process is introduced and that fresh observations are then made. In some cases it would be reasonable to expect a constant proportional change in stoppage rate, i.e. to expect that if the initial stoppage rates are $\lambda_1, \ldots, \lambda_k$ the final stoppage rates will be $\rho\lambda_1, \ldots, \rho\lambda_k$. If the observations are, in the initial period $(n_1, t_1), \ldots, (n_k, t_k)$, and in the final period $(n'_1, t'_1), \ldots, (n'_k, t'_k)$, we may define transformed variates $z_1, \ldots, z_k$ and $z'_1, \ldots, z'_k$ as in (12). Then $u_i = z'_i - z_i$ has mean $\mu = \log_{10}\rho$ and known variance $(\log_{10} e)^2 (v_i + v'_i) = (\log_{10} e)^2 w_i$ say. The best estimate of μ is thus

$$\hat{\mu} = \frac{\Sigma u_i w_i^{-1}}{\Sigma w_i^{-1}}, \quad \operatorname{var}\hat{\mu} = \frac{(\log_{10} e)^2}{\Sigma w_i^{-1}}. \tag{14}$$

An almost unbiased estimate of the parameter ρ is

$$\hat{\rho} = \left\{1 - \frac{1}{2\Sigma w_i^{-1}}\right\} 10^{\hat{\mu}}. \tag{15}$$

An approximate test of the hypothesis of a proportional change in stoppage rate may be obtained by calculating

$$\chi^2_{k-1} = \{\Sigma w_i^{-1}(u_i - \hat{\mu})^2\}(\log_{10} e)^2. \tag{16}$$

Example. The following data were the results of a sampling experiment with $\rho = 2$, the values of T_i, T'_i being chosen arbitrarily.

i	Initial		Final		u_i	$1/w_i$
	T_i (hr.)	n_i	T'_i (hr.)	n'_i		
1	30	2	28	5	0·428	1·429
2	26	4	41	21	0·522	3·360
3	21	2	24	3	0·118	1·200
4	45	18	16	11	0·235	6·828
5	30	4	32	12	0·449	3·000

$$\Sigma w_i^{-1} = 15{\cdot}817, \quad \Sigma u_i w_i^{-1} = 5{\cdot}4611, \quad \Sigma u_i^2 w_i^{-1} = 2{\cdot}1774.$$

The last two columns give

$$u_i = z'_i - z_i = \log_{10} T_i n'_i / T'_i n_i \quad \text{and} \quad w_i^{-1} = (v_i + v'_i)^{-1} = n_i n'_i/(n_i + n'_i),$$

no frequency being zero.

$$\hat{\mu} = \Sigma u_i w_i^{-1} / \Sigma w_i^{-1} = 0{\cdot}345 \quad \text{and} \quad \operatorname{var}\hat{\mu} = (0{\cdot}4343)^2/15{\cdot}817,$$

so that standard error $(\hat{\mu}) = 0{\cdot}109$. There is thus good agreement with the true value $\mu = \log_{10} 2 = 0{\cdot}3010$. ρ is estimated by

$$\hat{\rho} = \left\{1 - \frac{1}{2 \times 15{\cdot}817}\right\} 10^{\hat{\mu}}$$
$$= 2{\cdot}145.$$

The hypothesis of a proportional change in stoppage rate is tested by

$$\chi_4^2 = (2{\cdot}3026)^2 \Sigma w_i^{-1}(u_i - \hat{\mu})^2 = (2{\cdot}3026)^2 \{\Sigma w_i^{-1} u_i^2 - (\Sigma w_i^{-1})\, \hat{\mu}^2\}$$
$$= 1{\cdot}55, \text{ indicating a good agreement.}$$

This problem could be tackled without introducing the device (3) by the method of Dyke & Patterson (1952), using an iterative solution of maximum-likelihood equations. The present method is considerably quicker. The methods are probably asymptotically equivalent when the n_i and n'_i are all large. The transformation (12) can be applied in a similar way to more complicated problems involving Poisson variates.

I am grateful to Mr D. A. East and Miss Patricia Johnson who did most of the calculations.

REFERENCES

Barnard, G. A. (1946). *J.R. Statist. Soc. Suppl.* **8**, 1.
Barnard, G. A. (1947). *Biometrika*, **34**, 123.
Bartlett, M. S. & Kendall, D. G. (1946). *J.R. Statist. Soc. Suppl.* **8**, 128.
Chapman, D. G. (1952). *Ann. Inst. Statist. Math.* **4**, 45.
Dyke, G. V. & Patterson, H. D. (1952). *Biometrics*, **8**, 1.
Garwood, F. (1936). *Biometrika*, **28**, 437.
Maguire, B. A., Pearson, E. S. & Wynn, A. H. A. (1952). *Biometrika*, **39**, 168.
Przyborowski, J. & Wilenski, H. (1940). *Biometrika*, **31**, 313.
Wishart, J. (1947). *Biometrika*, **34**, 170.

ORTHOGONAL POLYNOMIAL FITTING

By JOHN WISHART AND THEOCHARIS METAKIDES

Statistical Laboratory, University of Cambridge

1. There are considerable advantages to be gained by using orthogonal polynomials when fitting curves to data by the method of least squares. The computer whose data are equally spaced, and of equal weight, is well catered for by tables, the best known of which are due to Fisher & Yates (1938–53), extended by Anderson & Houseman (1942); similar tables, due to Van der Reyden (1943), have been partly reproduced in the new *Biometrika Tables for Statisticians* (Pearson & Hartley, 1953). In the general case of unequally spaced and weighted data similar tables are impracticable; what the computer needs is a good calculation scheme, so designed that it makes little appeal to the memory, or the mathematics, and is as far as possible self-checking. There is only one mathematical basis for the methods, but a number of computing techniques have been devised, chiefly in order to carry out the inversion of the matrix whose elements are the equivalent of the variance and covariance estimates in the parallel problem of multiple regression. The only scheme known to the authors as having appeared in *Biometrika* is due to Isserlis (1927), but this needs revision. Elsewhere there are protagonists of the Gauss-Doolittle methods, described in more than one text-book; of the pivotal condensation method due to Aitken (1933, 1937); and of the Choleski method dealt with by Turing (1948), Fox, Huskey & Wilkinson (1948), Fox (1950), Fox & Hayes (1951), Rushton (1951), Hayes & Vickers (1951), and others. The recent papers by Guest (1950, 1953) deal with the mathematics of orthogonal polynomial fitting, and a practical computation scheme is given in the former, consisting of a combination of the Doolittle and pivotal condensation methods. (In fact, the available methods seem to differ only in the order and arrangement of the computations.)

The purpose of the present paper is to supplement the *Biometrika Tables* by offering a single recommended computation scheme for the general case. Like some other schemes, it is set out in tabular or 'blank' form, for ease of general manipulation. It is approached through multiple regression analysis, and to stress this fact the standard regression notation has been used. The arrangement is such that everything that can possibly be wanted is obtained within the tabular scheme. One point which may be considered convenient, e.g. for plotting purposes, is that while the orthogonal polynomials are calculated, the result is also given at any stage as a simple polynomial. One of the authors has examined the alternative schemes in application to the illustrative example given later in the paper, and the conclusion is that the method and scheme now offered are the most convenient for use by the average computer.

2. We begin by outlining a useful method of conducting a multiple regression analysis which brings to account one independent variate at a time, so that the successive terms of the final regression equation are orthogonal to one another, although it can in the end be converted to the usual form. Although well known, computational details of the method are not very readily accessible. They were published a few years ago by one of the authors (Wishart, 1950), while Woolf (1951) later gave practical directions for regression calculations by successive insertion of additional variates, using throughout the Gauss multipliers.

Consider a dependent variate x_n, usually assumed for significance test purposes to be normally distributed, and independent variates $x_1, x_2, \ldots, x_{n-1}$, usually assumed (for the same purposes) to be fixed; these are intended to be dealt with in the order stated. Let w denote the weight to be given to a sample observation. Add an additional 'independent' variate x_0 having a fixed value unity. Let Σ denote a sum of weighted products (or squares) of the designated variate measures taken over the sample, the size of which is N.

For $p, q = 0, 1, 2, \ldots, n$ we write

$$S_{pq} = \Sigma(wx_p x_q), \tag{2·1}$$

and then calculate in succession the partial sums

$$\left.\begin{aligned}
S_{p1\cdot0} &= S_{p1} - S_{p0}S_{10}/S_{00} && (p = 1, 2, \ldots, n),\\
S_{p2\cdot01} &= S_{p2\cdot0} - S_{p1\cdot0}S_{21\cdot0}/S_{11\cdot0} && (p = 2, 3, \ldots, n),\\
S_{p3\cdot012} &= S_{p3\cdot01} - S_{p2\cdot01}S_{32\cdot01}/S_{22\cdot01} && (p = 3, 4, \ldots, n),\\
S_{p4\cdot0123} &= S_{p4\cdot012} - S_{p3\cdot012}S_{43\cdot012}/S_{33\cdot012} && (p = 4, \ldots, n),
\end{aligned}\right\} \tag{2·2}$$

and so on. The displayed formulae (with $n = 4$) carry us as far as three independent variates, apart from x_0. We have in particular

$$S_{00} = \Sigma(w), \quad S_{p0} = \Sigma(wx_p).$$

The estimated regression coefficients, of progressively increasing order, are then calculated in turn, thus:

$$\left.\begin{aligned}
b_{n0} &= S_{n0}/S_{00},\\
b_{n1\cdot0} &= S_{n1\cdot0}/S_{11\cdot0},\\
b_{n2\cdot01} &= S_{n2\cdot01}/S_{22\cdot01},\\
b_{n3\cdot012} &= S_{n3\cdot012}/S_{33\cdot012},
\end{aligned}\right\} \tag{2·3}$$

and so on. In particular $b_{n0} = \Sigma(wx_n)/\Sigma(w) = \bar{x}_n.$

Partial x-variates form the next sequence, and are calculated as below:

$$\left.\begin{aligned}
x_{p\cdot0} &= x_p - S_{p0}x_0/S_{00},\\
x_{p\cdot01} &= x_{p\cdot0} - S_{p1\cdot0}x_{1\cdot0}/S_{11\cdot0},\\
x_{p\cdot012} &= x_{p\cdot01} - S_{p2\cdot01}x_{2\cdot01}/S_{22\cdot01},
\end{aligned}\right\} \tag{2·4}$$

and so on. The first of these is $x_{p\cdot0} = x_p - \bar{x}_p$.

The multiple regression equation, written as a sum of orthogonal components, is then

$$X_n = b_{n0}x_0 + b_{n1\cdot0}x_{1\cdot0} + b_{n2\cdot01}x_{2\cdot01} + b_{n3\cdot012}x_{3\cdot012} + \ldots, \tag{2·5}$$

of which the first term on the right-hand side is $\bar{x}_n$. For each term fitted there is a 'sum of squares' due to the regression, with one degree of freedom and, when all have been fitted to the chosen number of independent variates, there is a residual sum of squares. The analysis of variance, taken up to three variates, is shown in Table 1; in this, on the usual assumptions, the various mean squares are distributed as variance estimates independently of one another, with the assigned numbers of degrees of freedom. In the table the contribution due to $\bar{x}_n$ has been kept separate from S_{nn}; in the more usual form the total sum of squares would be $S_{nn\cdot0}$, with $N-1$ degrees of freedom. The overall variance-ratio test of regression after bringing in three variates is illustrated, but note that this test could have been applied at an earlier stage, also that at any stage the mean square due to a single regression coefficient is tested against the residual mean square available at that stage.

Table 1

Regression coefficient	Degrees of freedom	Sums of squares		Mean squares
b_{n0}	1	$b_{n0}S_{n0} = S_{nn}-S_{nn\cdot 0}$		
$b_{n1\cdot 0}$	1	$b_{n1\cdot 0}S_{n1\cdot 0} = S_{nn\cdot 0}-S_{nn\cdot 01}$	$S_{nn\cdot 0}-S_{nn\cdot 0123}$	V_b
$b_{n2\cdot 01}$	1	$b_{n2\cdot 01}S_{n2\cdot 01} = S_{nn\cdot 01}-S_{nn\cdot 012}$		
$b_{n3\cdot 012}$	1	$b_{n3\cdot 012}S_{n3\cdot 012} = S_{nn\cdot 012}-S_{nn\cdot 0123}$		
Residual	$N-4$	$S_{nn\cdot 0123}$		V_r
Total	N	S_{nn}		

In practice it is helpful to choose as x_1 that variate which is most nearly expected to be associated with x_n; then as x_2 the next in expected strength of association, and so on. The calculations can be reviewed when a fitted term is found to be insignificant; the appropriate variate may be discarded and another substituted. As a result of this procedure it may be possible to say with a minimum amount of labour that x_n is adequately 'predicted' by a limited number only of independent variates, a matter of some importance in a variety of problems, for example, in physical anthropology where large numbers of body and other measurements may be available, and for practical reasons it is desirable to deal with a small number of variates only for purposes of description.

The more symmetrical arrangement of the regression equation, giving all the partial regression coefficients of the highest order, and also tests of their individual significance and a final check, require additional calculations which can be readily inferred from the application of the method which is made in §3 to orthogonal polynomial fitting.

It will be noted that the first stage in the scheme involves the determination of the weighted sums of products (and squares) of deviations of variate measures from their (weighted) means. That is, $S_{pq\cdot 0}$ is determined from S_{pq} by removing the usual 'correction factor' $S_{p0}S_{q0}/S_{00}$. If the quantities $S_{pq\cdot 0}$ are given in the first instance, one stage in the computations is eliminated.

3. Turning now to the problem of orthogonal polynomial fitting, we read x_n as y, to distinguish the dependent variate. We then let

$$x_0 = x^0 = 1, \quad x_1 = x^1 = x, \quad x_2 = x^2, \quad \dots,$$

so that the independent variates become the powers of x, including the dummy variate. In the special case of $n = N$, the sample size, so that there are $N-1$ independent variates, the multiple regression equation will then become the polynomial of degree $N-1$ of perfect fit. The usual case is, however, where we endeavour to fit a polynomial of degree p, where $p < N-1$. We then have

$$S_{r+s,0} = S_{r+s-1,1} = \dots = S_{rs} = \dots = S_{0,r+s} = \Sigma(wx^{r+s}), \tag{3·1}$$

where r and s take independently the values $0, 1, 2, \dots, p$. We define also

$$S_{yr} = \Sigma(wyx^r), \text{ needed for } r = 0, 1, 2, \dots, p. \tag{3·2}$$

Equations (2·3) will be needed for $n = 1, 2, \dots, p$ as well as for $n = y$, and we shall also find it convenient to calculate the quantities $b_{20\cdot 1}, b_{30\cdot 12}, \dots$ which are defined below. In particular b_{r0} is the mean value of the rth powers of the x's.

In virtue of (2·3), equations (2·2) may be calculated as

$$\left.\begin{aligned} S_{r1\cdot 0} &= S_{r1}-b_{r0}S_{10} & &(r = 1, 2, \dots, p \text{ and } y),\\ S_{r2\cdot 01} &= S_{r2\cdot 0}-b_{r1\cdot 0}S_{21\cdot 0} & &(r = 2, 3, \dots, p \text{ and } y),\\ S_{r3\cdot 012} &= S_{r3\cdot 01}-b_{r2\cdot 01}S_{32\cdot 01} & &(r = 3, \dots, p \text{ and } y),\end{aligned}\right\} \tag{3·3}$$

and so on. Equation (2·5) becomes, in the present case,

$$Y = b_{y0}x_0 + b_{y1\cdot 0}x_{1\cdot 0} + b_{y2\cdot 01}x_{2\cdot 01} + b_{y3\cdot 012}x_{3\cdot 012} + \dots. \tag{3·4}$$

In Fisher's notation (Fisher, 1925) the successive orthogonal polynomials are $\xi_0 = 1$, ξ_1, $\xi_2, \dots$, with multipliers $A, B, C, \dots$, so that the fitted polynomial is

$$Y = A + B\xi_1 + C\xi_2 + D\xi_3 + \dots. \tag{3·5}$$

From equations (2·4), by inserting b-coefficients, we have

$$\left.\begin{aligned} x_{1\cdot 0} &= x_1 - b_{10}, \\ x_{2\cdot 01} &= x_{2\cdot 0} - b_{21\cdot 0}x_{1\cdot 0}, \\ x_{3\cdot 012} &= x_{3\cdot 01} - b_{32\cdot 01}x_{2\cdot 01}, \end{aligned}\right\} \tag{3·6}$$

and so on. In virtue of the relations

$$\bar{x} = b_{10}, \quad \overline{x^2} = b_{20}, \quad \overline{x^3} = b_{30}, \dots;$$

$$\left.\begin{aligned} b_{20} - b_{21\cdot 0}b_{10} &= b_{20\cdot 1}; \\ b_{30} - b_{31\cdot 0}b_{10} - b_{32\cdot 01}b_{20\cdot 1} &= b_{30\cdot 12}; \\ b_{31\cdot 0} - b_{32\cdot 01}b_{21\cdot 0} &= b_{31\cdot 02}, \end{aligned}\right\} \tag{3·7}$$

and so on, we find, after appropriate substitution and simplification, replacing $x_0, x_1, x_2, \cdots$ by $x^0, x^1, x^2, \dots$, that

$$\left.\begin{aligned} x_0 &= \xi_0 = 1, \\ x_{1\cdot 0} &= \xi_1 = -b_{10} + x, \\ x_{2\cdot 01} &= \xi_2 = -b_{20\cdot 1} - b_{21\cdot 0}x + x^2, \\ x_{3\cdot 012} &= \xi_3 = -b_{30\cdot 12} - b_{31\cdot 02}x - b_{32\cdot 01}x^2 + x^3, \end{aligned}\right\} \tag{3·8}$$

and so on, and that

$$A = b_{y0}, \quad B = b_{y1\cdot 0}, \quad C = b_{y2\cdot 01}, \quad D = b_{y3\cdot 012}, \dots. \tag{3·9}$$

By combining we may write down the different terms of Y separately, or they may be cumulated one at a time, thus having the appropriate fitted curve at any stage expressed as a simple polynomial in x, as follows:

$$\left.\begin{aligned} Y_0 &= b_{y0}, \\ Y_1 &= b_{y0\cdot 1} + b_{y1\cdot 0}x, \\ Y_2 &= b_{y0\cdot 12} + b_{y1\cdot 02}x + b_{y2\cdot 01}x^2, \\ Y_3 &= b_{y0\cdot 123} + b_{y1\cdot 023}x + b_{y2\cdot 013}x^2 + b_{y3\cdot 012}x^3, \end{aligned}\right\} \tag{3·10}$$

and so on, where Y_r denotes (3·4) taken to degree r. The coefficients are calculated from (2·3), putting $n = y$, and from (3·7) with y written for $2, 3, \dots$.

The analysis of variance will be as in Table 1 with y substituted for n. The estimated variances of the coefficients of the power series in (3·10) may be written down in terms of the following 'Gauss multipliers':

$$\left.\begin{aligned} &c_{00} = S_{00}^{-1}, \\ &c_{00\cdot 1} = c_{00} - b_{10}c_{10}, & &c_{11\cdot 0} = S_{11\cdot 0}^{-1}, \\ &c_{00\cdot 12} = c_{00\cdot 1} - b_{20\cdot 1}c_{20\cdot 1}, & &c_{11\cdot 02} = c_{11\cdot 0} - b_{21\cdot 0}c_{21\cdot 0}, \\ &c_{00\cdot 123} = c_{00\cdot 12} - b_{30\cdot 12}c_{30\cdot 12}, & &c_{11\cdot 023} = c_{11\cdot 02} - b_{31\cdot 02}c_{31\cdot 02}, \\ &c_{22\cdot 01} = S_{22\cdot 01}^{-1}, \\ &c_{22\cdot 013} = c_{22\cdot 01} - b_{32\cdot 01}c_{32\cdot 01}, & &c_{33\cdot 012} = S_{33\cdot 012}^{-1}, \end{aligned}\right\} \tag{3·11}$$

in which

$$c_{10} = -b_{10}c_{11\cdot 0},\quad c_{20\cdot 1} = -b_{20\cdot 1}c_{22\cdot 01},\quad c_{21\cdot 0} = -b_{21\cdot 0}c_{22\cdot 01},\quad c_{30\cdot 12} = -b_{30\cdot 12}c_{33\cdot 012},$$
$$c_{31\cdot 02} = -b_{31\cdot 02}c_{33\cdot 012},\quad c_{32\cdot 01} = -b_{32\cdot 01}c_{33\cdot 012},$$

and so on. If we denote the 'mean squares' by

$$S_{yy\cdot 0}/(N-1) = s_0^2,\quad S_{yy\cdot 01}/(N-2) = s_1^2, \ldots,$$

then, V denoting 'estimated variance of', we have

$$\left.\begin{array}{llll} V(b_{y0}) & = c_{00}s_0^2; & & \\ V(b_{y0\cdot 1}) & = c_{00\cdot 1}s_1^2, & V(b_{y1\cdot 0}) & = c_{11\cdot 0}s_1^2, \\ V(b_{y0\cdot 12}) & = c_{00\cdot 12}s_2^2, & V(b_{y1\cdot 02}) & = c_{11\cdot 02}s_2^2, \\ V(b_{y0\cdot 123}) & = c_{00\cdot 123}s_3^2, & V(b_{y1\cdot 023}) & = c_{11\cdot 023}s_3^2, \\ V(b_{y2\cdot 01}) & = c_{22\cdot 01}s_2^2; & & \\ V(b_{y2\cdot 013}) & = c_{22\cdot 013}s_3^2, & V(b_{y3\cdot 012}) & = c_{33\cdot 012}s_3^2, \end{array}\right\} \tag{3·12}$$

and so on. Finally, it may be useful to have formulae for the estimated variances of the ordinates of the different order curves. Calling these ordinates $\hat{Y}_0, \hat{Y}_1, \ldots$, we have

$$\left.\begin{array}{l} V(\hat{Y}_0) = c_{00}\xi_0^2 s_0^2, \\ V(\hat{Y}_1) = (c_{00}\xi_0^2 + c_{11\cdot 0}\xi_1^2)\, s_1^2, \\ V(\hat{Y}_2) = (c_{00}\xi_0^2 + c_{11\cdot 0}\xi_1^2 + c_{22\cdot 01}\xi_2^2)\, s_2^2, \\ V(\hat{Y}_3) = (c_{00}\xi_0^2 + c_{11\cdot 0}\xi_1^2 + c_{22\cdot 01}\xi_2^2 + c_{33\cdot 012}\xi_3^2)\, s_3^2, \end{array}\right\} \tag{3·13}$$

and so on, in which, of course, $\xi_0 = 1$ and the values of $\xi_1, \xi_2, \ldots$ are obtained from (3·8) for any value of x.

4. For practical computation purposes it is desirable to systematize the evaluation of the quantities set out in succession in § 3. This leads to the tabular scheme set out in Table 2, which proceeds as far as the fitting of a cubic, and can readily be extended for higher degree polynomials. Owing to the symmetry of the original matrix of sums of squares and products, one need only enter quantities on and to the right of the diagonals. This is indicated by the heavy vertical lines in the upper part of the table. A small number of supplementary lines are needed to obtain the actual polynomial coefficients; these are on the left immediately below the main table. The column on the right is used for the various contributions to the sum of squares, and from this the analysis of variance is readily written down. The check column is of the usual character for this type of work. The operations of the rest of the table are followed here, and then in those lines where a tick mark is shown the check is that the number in this column is the sum of all numbers in that line to the left, stopping at the heavy vertical rule. A useful final check is shown in the lower box on the right. The table shown in lines 1′–16′ is a subsidiary one, giving the matrix of the c's, inverse to that of the S's. Only the diagonal c's are required for the estimated variances of (3·12) and (3·13), but the whole matrix has been included for the sake of completeness. It is useful to have all the c's readily available, particularly when the scheme is used for multiple regression (see § 2), since they are needed in calculating adjusted coefficients of the surviving independent variates, without redoing the calculations, when it has been decided to omit any such variate. The appropriate formulae for this purpose were given by R. A. Fisher in the 5th edition of *Statistical Methods for Research Workers*, and discussed by Cochran (1938).

Table 2

						Check column	Sums of squares
1	S_{00}	S_{10}	S_{20}	S_{30}	S_{y0}		S_{yy}
2	-1	$-b_{10}$	$-b_{20}$	$-b_{30}$	$-b_{y0}$	✓	$b_{y0}S_{y0}$
3		S_{11}	S_{21}	S_{31}	S_{y1}		
4	$-b_{10}$	$-b_{10}S_{10}$	$-b_{10}S_{20}$	$-b_{10}S_{30}$	$-b_{10}S_{y0}$		
5	$-b_{10}$	$S_{11\cdot 0}$	$S_{21\cdot 0}$	$S_{31\cdot 0}$	$S_{y1\cdot 0}$	✓	
6	c_{10}	-1	$-b_{21\cdot 0}$	$-b_{31\cdot 0}$	$-b_{y1\cdot 0}$	✓	$b_{y1\cdot 0}S_{y1\cdot 0}$
7			S_{22}	S_{32}	S_{y2}		
8	$-b_{20}$		$-b_{20}S_{20}$	$-b_{20}S_{30}$	$-b_{20}S_{y0}$		
9	$b_{21\cdot 0}b_{10}$	$-b_{21\cdot 0}$	$-b_{21\cdot 0}S_{21\cdot 0}$	$-b_{21\cdot 0}S_{31\cdot 0}$	$-b_{21\cdot 0}S_{y1\cdot 0}$		
10	$-b_{20\cdot 1}$	$-b_{21\cdot 0}$	$S_{22\cdot 01}$	$S_{32\cdot 01}$	$S_{y2\cdot 01}$	✓	
11	$c_{20\cdot 1}$	$c_{21\cdot 0}$	-1	$-b_{32\cdot 01}$	$-b_{y2\cdot 01}$	✓	$b_{y2\cdot 01}S_{y2\cdot 01}$
12				S_{33}	S_{y3}		
13	$-b_{30}$			$-b_{30}S_{30}$	$-b_{30}S_{y0}$		
14	$b_{31\cdot 0}b_{10}$	$-b_{31\cdot 0}$		$-b_{31\cdot 0}S_{31\cdot 0}$	$-b_{31\cdot 0}S_{y1\cdot 0}$		
15	$b_{32\cdot 01}b_{20\cdot 1}$	$b_{32\cdot 01}b_{21\cdot 0}$	$-b_{32\cdot 01}$	$-b_{32\cdot 01}S_{32\cdot 01}$	$-b_{32\cdot 01}S_{y2\cdot 01}$		
16	$-b_{30\cdot 12}$	$-b_{31\cdot 02}$	$-b_{32\cdot 01}$	$S_{33\cdot 012}$	$S_{y3\cdot 012}$	✓	
17	$c_{30\cdot 12}$	$c_{31\cdot 02}$	$c_{32\cdot 01}$	-1	$-b_{y3\cdot 012}$	✓	$b_{y3\cdot 012}S_{y3\cdot 012}$
18	b_{y0}						
19	$-b_{10}b_{y1\cdot 0}$	$b_{y1\cdot 0}$					$S_{yy}-S_{yy\cdot 0123}$
20	$b_{y0\cdot 1}$	$b_{y1\cdot 0}$					
21	$-b_{20\cdot 1}b_{y2\cdot 01}$	$-b_{21\cdot 0}b_{y2\cdot 01}$	$b_{y2\cdot 01}$				
22	$b_{y0\cdot 12}$	$b_{y1\cdot 02}$	$b_{y2\cdot 01}$				
23	$-b_{30\cdot 12}b_{y3\cdot 012}$	$-b_{31\cdot 02}b_{y3\cdot 012}$	$-b_{32\cdot 01}b_{y3\cdot 012}$	$b_{y3\cdot 012}$			
24	$b_{y0\cdot 123}$	$b_{y1\cdot 023}$	$b_{y2\cdot 013}$	$b_{y3\cdot 012}$			

$$
\begin{aligned}
&b_{y0\cdot 123}S_{y0}\\
&+b_{y1\cdot 023}S_{y1}\\
&+b_{y2\cdot 013}S_{y2}\\
&+b_{y3\cdot 012}S_{y3}\\
&=S_{yy}-S_{yy\cdot 0123}
\end{aligned}
$$

Final check

1′	c_{00}			
2′	$-b_{10}c_{10}$			
3′	$c_{00\cdot 1}$			
4′	$-b_{20\cdot 1}c_{20\cdot 1}$			
5′	$c_{00\cdot 12}$			
6′	$-b_{30\cdot 12}c_{30\cdot 12}$			
7′	$c_{00\cdot 123}$			
8′	c_{10}	$c_{11\cdot 0}$		
9′	$-b_{21\cdot 0}c_{20\cdot 1}$	$-b_{21\cdot 0}c_{21\cdot 0}$		
10′	$c_{10\cdot 2}$	$c_{11\cdot 02}$		
11′	$-b_{31\cdot 02}c_{30\cdot 12}$	$-b_{31\cdot 02}c_{31\cdot 02}$		
12′	$c_{10\cdot 23}$	$c_{11\cdot 023}$		
13′	$c_{20\cdot 1}$	$c_{21\cdot 0}$	$c_{22\cdot 01}$	
14′	$-b_{32\cdot 01}c_{30\cdot 12}$	$-b_{32\cdot 01}c_{31\cdot 02}$	$-b_{32\cdot 01}c_{32\cdot 01}$	
15′	$c_{20\cdot 13}$	$c_{21\cdot 03}$	$c_{22\cdot 013}$	
16′	$c_{30\cdot 12}$	$c_{31\cdot 02}$	$c_{32\cdot 01}$	$c_{33\cdot 012}$

Subsidiary table giving the *c*'s.

Table 3

1	14	91	803	7705	78	8691	538
2	−1	−6·5	−57·357143	−550·357143	−5·571429	−620·78571 ✓	434·5715
3		803	7705	77447	589	86635	
4	−6·5	−591·5	−5219·5	−50082·5	−507	−56491·5	
5	−6·5	211·5	2485·5	27364·5	82	30143·5 ✓	
6	−0·0307329	−1	−11·751773	−129·382979	−0·387707	−142·52246 ✓	31·7920
7			77447	804121	5049	895125	
8	−57·357143		−46057·7858	−441936·787	−4473·8573	−498490·930	
9	76·386525	−11·751773	−29209·0318	−321581·392	−963·6454	−354239·569	
10	19·029382	−11·751773	2180·1824	40602·821	−388·5027	42394·501 ✓	
11	0·00872834	−0·00539027	−1	−18·623589	0·178197	−19·44539 ✓	69·2300
12				8555663	46207	9491143	
13	−550·357143			−4240501·8	−42927·857	−4783153·93	
14	840·989364	−129·382979		−3540500·5	−10609·404	−3900055·83	
15	−354·395389	218·860190	−18·623589	−756170·3	7235·315	−789537·76	
16	−63·763168	89·477211	−18·623589	18490·4	−94·946	18395·48 ✓	
17	−0·00344845	0·00483911	−0·00100720	−1	0·005135	−0·994865 ✓	0·4875
18	5·571429						536·0810
19	−2·520096	0·387707					
20	3·051333	0·387707					
21	−3·390979	2·094131	−0·178197				−0·9533 1191·1777
22	−0·339646	2·481838	−0·178197				−416·8707
23	0·327424	−0·459465	0·095632	−0·005135			−237·2729
24	−0·012222	2·022373	−0·082565	−0·005135			536·0808

1′	0·0714286			
2′	0·1997638			
3′	0·2711924			
4′	0·1660949			
5′	0·4372873			
6′	0·2198841			
7′	0·6571714			
8′	−0·0307329	0·0047281		
9′	−0·1025736	0·0633452		
10′	−0·1333065	0·0680733		
11′	−0·3085577	0·4329902		
12′	−0·4418642	0·5010635		
13′	0·00872834	−0·00539027	0·000458677	
14′	0·06422252	−0·09012160	0·018757678	
15′	0·07295086	−0·09551187	0·019216355	
16′	−0·00344845	0·00483911	−0·00100720	0·0000540821

Little explanation is needed of the operation of this scheme. The primary S's (from (3·1) and (3·2)) are entered in lines 1, 3, 7 and 12. To form line 2, all quantities in line 1 are divided by S_{00}, whose reciprocal is c_{00}, a quantity which is at the same time entered on line 1′. In this and subsequent calculations the signs are changed to the right of the heavy vertical rule. Line 6 is similarly obtained from line 5, line 11 from line 10 and line 17 from line 16, by dividing by $S_{11\cdot 0}$, $S_{22\cdot 01}$ and $S_{33\cdot 012}$, whose reciprocals are $c_{11\cdot 0}$, $c_{22\cdot 01}$ and $c_{33\cdot 012}$ respectively. Line 4 comes from line 1 as shown using the multiplier $-b_{10}$; the next line is the sum of lines 3 and 4, as is indicated by the short horizontal rules. At the next stage lines 8 and 9 are derived from lines 1 and 5 by the use of b-multipliers already determined in lines 2 and 6; line 10 is another summation, this time of lines 7, 8 and 9. And so on. The appropriate b's and c's are entered in the supplementary lines and subsidiary table as calculated, the subsequent calculations in these lines being as shown; vertical additions are made in pairs of lines, for example line 20 is the sum of lines 18 and 19, line 3′ is the sum of lines 1′ and 2′, and so on.

We are only concerned in this paper with the fitting of orthogonal polynomials, and the computation scheme has therefore been drawn up with this end in view. But it should be pointed out that the scheme generally is one for matrix inversion, at any rate where the simultaneous equations whose solution is desired are symmetrical, and it may be found to compare favourably with other schemes which have been advocated in the past. It requires no adaptation to be used in standard multiple regression work, as described in §2.

5. *Numerical example*. We shall fit as far as the third degree polynomial the following data:

Weight (w)	1	2	1	1	2	3	1	2	1
Values of x	0	1	3	4	7	8	10	11	12
Values of y	0	2	5	6	9	8	7	5	4

Since for a polynomial of the pth degree we require the weighted sums of the first $2p$ powers of x and the weighted sums of products of y with the first p powers of x, as well as the weighted sum of squares of y (for the analysis of variance), the first step consists in calculating these quantities, either by means of a tabular build-up or by using tables of powers. The calculation yields

$$\begin{array}{llrl}
S_{00} & = \Sigma(w) \;= & 14 & S_{y0} = \Sigma(wy) \;= \;\; 78 \\
S_{10} = S_{01} & = \Sigma(wx) \;= & 91 & S_{y1} = \Sigma(wyx) \;= \;\; 589 \\
S_{20} = S_{11} = S_{02} & = \Sigma(wx^2) = & 803 & S_{y2} = \Sigma(wyx^2) = \;\; 5049 \\
S_{30} = S_{21} = S_{12} = S_{03} & = \Sigma(wx^3) = & 7705 & S_{y3} = \Sigma(wyx^3) = 46207 \\
S_{31} = S_{22} = S_{13} & = \Sigma(wx^4) = & 77447 & \\
S_{32} = S_{23} & = \Sigma(wx^5) = & 804121 & \\
S_{33} & = \Sigma(wx^6) = & 8555663 & S_{yy} = \Sigma(wy^2) \;= \;\; 538
\end{array}$$

These quantities are now entered in the appropriate places of Table 2, and calculation proceeds. This is shown in Table 3, in which enough decimal places have been retained to show agreement at the final check to two or more places of decimals. Fewer figures need be written down if the accuracy desired is not of this order. For all ordinary purposes a maximum of four decimal places, or perhaps four significant figures, will suffice.

The fitted curves of degrees 0, 1, 2 and 3 are:

$$Y_0 = 5\cdot57143,$$
$$Y_1 = 3\cdot05133+0\cdot38771x,$$
$$Y_2 = -0\cdot33965+2\cdot48184x-0\cdot17820x^2,$$
$$Y_3 = -0\cdot01222+2\cdot02237x-0\cdot08256x^2-0\cdot00514x^3.$$

The analysis of variance need not be separately written out. The residual sum of squares, from Table 3, is $538-536\cdot0810 = 1\cdot9190$, with 5 degrees of freedom. s_3^2 is therefore 0·3838, whence it follows that the linear and quadratic terms, with mean squares 31·7920 and 69·2300, are significant at the 0·1 % point. The cubic term, with mean square 0·4875, is not significant. If this were a practical problem we should either go on to one more stage, or, if that was not thought worth while, rest content with describing the data by means of the above quadratic Y_2. This may be plotted against the actual data by substituting for x, when we get the following results:

Weight (w)	1	2	1	1	2	3	1	2	1
Values of y	0	2	5	6	9	8	7	5	4
Values of Y_2	−0·34	1·96	5·50	6·74	8·30	8·11	6·66	5·40	3·78

s_2^2 will be $(0\cdot4875+1\cdot9190)/6 = 2\cdot4065/6 = 0\cdot4011$. This is multiplied by $c_{00\cdot12}$, $c_{11\cdot02}$ and $c_{22\cdot01}$ respectively, and the square roots taken, to deduce the standard errors of the coefficients $b_{y0\cdot12}$, $b_{y1\cdot02}$ and $b_{y2\cdot01}$ respectively, of Y_2. These standard errors are 0·4188, 0·1652 and 0·0136 respectively. From these it appears that $b_{y1\cdot02}$ and $b_{y2\cdot01}$ are large in comparison with their standard errors, and very significant, but that $b_{y0\cdot12}$ is not significantly different from zero. The calculation of the estimated standard error of an estimated Y_2 may be illustrated by taking the value 6·74, which corresponds to $x = 4$. $s_2^2 = 0\cdot4011$ has to be multiplied, as in (3·13), by

$$c_{00}\xi_0^2+c_{11\cdot0}\xi_1^2+c_{22\cdot01}\xi_2^2,$$

in which the c's are obtained from Table 3 and the ξ's are given in (3·8) in terms of the b's, also obtained from Table 3, and x. Putting $x = 4$, multiplying the above expression by 0·4011 and extracting the square root, we find the estimated standard error at this point to be 0·26.

REFERENCES

AITKEN, A. C. (1933). *Proc. Roy. Soc. Edinb.* **53**, 54.
AITKEN, A. C. (1937). *Proc. Roy. Soc. Edinb.* **57**, 172.
ANDERSON, R. L. & HOUSEMAN, E. E. (1942). *Res. Bull. no.* 297, Ames, Iowa.
COCHRAN, W. G. (1938). *J.R. Statist. Soc. Suppl.* **5**, 171.
FISHER, R. A. (1925–50). *Statistical Methods for Research Workers.* Oliver and Boyd.
FISHER, R. A. & YATES, F. (1938–53). *Statistical Tables.* Oliver and Boyd.
FOX, L. (1950). *J.R. Statist. Soc.* B, **12**, 120.
FOX, L. & HAYES, J. G. (1951). *J.R. Statist. Soc.* B, **13**, 83.
FOX, L., HUSKEY, H. D. & WILKINSON, J. H. (1948). *Quart. J. Mech. Appl. Math.* **1**, 149.
GUEST, P. G. (1950). *Phil. Mag.* [7], **41**, 124.
GUEST, P. G. (1953). *Biometrika*, **40**, 229.
HAYES, J. G. & VICKERS, T. (1951). *Phil. Mag.* [7], **42**, 1387.
ISSERLIS, L. (1927). *Biometrika*, **19**, 87.
PEARSON, E. S. & HARTLEY, H. O. (1953). *Biometrika Tables for Statisticians.* Cambridge University Press.
RUSHTON, S. (1951). *J.R. Statist. Soc.* B, **13**, 92.
TURING, A. M. (1948). *Quart. J. Mech. Appl. Math.* **1**, 287.
VAN DER REYDEN, D. (1943). *Onderstepoort J. Vet. Sci.* **18**, 355.
WISHART, J. (1950). *Field Trials. II. The Analysis of Covariance.* Tech. Comm. no. 15 (Commonwealth Bureau of Plant Breeding and Genetics, Cambridge).
WOOLF, B. (1951). *J.R. Statist. Soc.* B, **13**, 100.

POPULATION DIFFERENCES BETWEEN SPECIES GROWING ACCORDING TO SIMPLE BIRTH AND DEATH PROCESSES

By J. H. DARWIN

1. Introduction

In some ecological problems, treated especially by C. B. Williams (e.g. 1944), the numerical sizes of different species of plants or animals present in a sampling region have followed Fisher's logarithmic law, viz. the expected number of species having an observed population size n is $\alpha x^n/n$ $(n = 1, 2, \ldots)$. In this α is a constant depending only on the biological association being sampled and not on the interval of sampling. This law for the observed species was derived from the negative binomial distribution, for which the probability that a species has n members is given by

$$P_n = \frac{\Gamma(n+k)}{\Gamma(n+1)\,\Gamma(k)}\frac{\sigma^n}{(1+\sigma)^{n+k}} \quad (n = 0, 1, \ldots),$$

by allowing the parameter k to tend to 0, and the number N of possible species to tend to infinity in such a way that $Nk = \alpha$, $\sigma/(1+\sigma) = x$. The negative binomial is a compound Poisson distribution, i.e. it arises from the assumption that the population of each species is represented by a Poisson distribution with mean m, the different m's for the different species being regarded as a random sample from the distribution

$$(1/\sigma^k\Gamma(k))\,x^{k-1}\,e^{-x/\sigma}\,dx.$$

Thus if there is a large number of species, mostly rare, in danger of losing a member to the sample, Fisher's logarithmic law might be a reasonable description of the heterogeneity between them.

One may further discuss this heterogeneity in terms of the manner in which different species grow in the sampling region. For instance, D. G. Kendall (1948*a*) considered a population expanding in a simple birth and death process in which the probabilities of a birth or a death from among n members alive at time t are, for the next interval dt, $n\lambda\,dt$ and $n\mu\,dt$ respectively; in dt there is also a probability $\kappa\,dt$ of the entry into the region of a new member from outside; λ, μ and κ are constants. These transition probabilities lead to the negative binomial as a current description of the size of a species whose first member in the region is an immigrant. When the immigration constant κ is small compared with λ and μ, a logarithmic law with a zero class results. The often observed logarithmic type differences in specific population size would then be reasonable if the species come into the region at times randomly scattered in the interval $(0, t)$, and grow roughly according to the same geometric birth and death process. The possibilities of such 'explanations' of heterogeneity between species can be formally investigated for different theoretical distributions. It will be shown that when the population size is bounded, the form of solution of processes for which the transition probability rates do not vary with time enables one to test if a particular distribution can describe the state of the process at any time. When the transition probability rates do vary with time, it appears that there may be many stochastic geneses of such well-known distributions as the negative binomial and the Poisson. Finally, a simple evolutionary model for the changing number of species is applied to species which expand geometrically, and logarithmic type distributions of the observed population sizes of the species again occur.

2. FINITE DISTRIBUTIONS

If the size of the population can take any of the values 0, 1, ..., N, the equations of the birth and death process are

$$\left.\begin{aligned} \frac{dP_n}{dt} &= \alpha_{n+1}P_{n+1} - (\alpha_n+\beta_n)P_n + \beta_{n-1}P_{n-1} \quad (n = 1, 2, \ldots, N-1), \\ \frac{dP_0}{dt} &= \alpha_1 P_1 - \beta_0 P_0, \quad \frac{dP_N}{dt} = -\alpha_N P_N + \beta_{N-1}P_{N-1}, \end{aligned}\right\} \tag{1}$$

in which P_n is the probability that a species has size n (n could be the number of regions out of a possible N in which the species could be found); α_n is the probability death-rate and β_n the probability birth-rate when the size is n.

2·1. *Transition probability rates not varying with time.* When α_n and β_n are constants, which is, for instance, reasonable when the spores or seeds of a plant or the offspring of an animal can travel over distances of the same order as the dimensions of a region, the set of equations can be written as

$$\frac{d\mathbf{P}}{dt} = -A\mathbf{P}, \tag{2}$$

where $\mathbf{P}$ is the column vector of probabilities $P_0, P_1, \ldots, P_N$, and A is the matrix with non-zero elements only in the three main diagonals

$$\begin{pmatrix} \beta_0 & -\alpha_1 & & & \\ -\beta_0 & \alpha_1+\beta_1 & -\alpha_2 & & \\ & -\beta_1 & \ddots & & \\ & & & \alpha_{N-1}+\beta_{N-1} & -\alpha_N \\ & & & -\beta_{N-1} & \alpha_N \end{pmatrix}.$$

It is well known that if the latent roots λ_i $(i = 0, 1, \ldots, N)$ of this matrix are distinct, the solution P_n is of the form $\sum_{i=0}^{N} A_{ni}\, e^{-\lambda_i t}$, where the A_{ni} are constants independent of time and embracing the initial conditions. It is further possible to show that if $\alpha_1, \ldots, \alpha_N, \beta_1, \ldots, \beta_{N-1}$ are > 0, these roots λ_i are real, distinct and positive (a result also established by B. J. Prendiville, unpublished). This follows from an expansion of the leading $(n+1)\times(n+1)$ minor of $|\lambda I - A|$ as $D_n(\lambda)$ in the form

$$D_n(\lambda) = (\lambda - (\alpha_n+\beta_n))\, D_{n-1}(\lambda) - \alpha_n\beta_{n-1}D_{n-2}(\lambda). \tag{3}$$

An induction hypothesis that $D_{n-1}(\lambda)$ and $D_{n-2}(\lambda)$ have real distinct roots $\mu_0, \mu_1, \ldots, \mu_{n-1}$; $\nu_0, \nu_1, \ldots, \nu_{n-2}$, such that $\mu_0 < \nu_0 < \mu_1 < \ldots < \nu_{n-2} < \mu_{n-1}$ leads to the conclusion that $D_n(\lambda)$ has real distinct roots separating those of $D_{n-1}(\lambda)$. An expansion of $|\lambda I - A|$ shows, by the sign rule, that there can be no real negative roots, and that the lowest root of $D_n(\lambda)$ for $n < N$ is > 0. The result then follows. It can be used as a criterion for testing if possible processes of this simple type could lead to given distributions.

Thus, if

$$\frac{dP_n}{dt} = P_n\, g_n(t) \quad \text{and} \quad P_{n+1} = r(t)\, P_n,$$

(1) becomes, when $P(t)$ is not identically zero,

$$g_n(t) = \alpha_{n+1} r_n - (\alpha_n + \beta_n) + \beta_{n-1}/r_{n-1}. \tag{4}$$

This is a set of $N+1$ equations in the $N+1$ coefficients of $e^{-\lambda_i t}$.

2·2. *Examples.* In the binomial, $P_n = \binom{N}{n} p^n (1-p)^{N-n}$, since $P_N = p^N$ must be of the form $\sum_{i=0}^{N} A_{ni} e^{-\lambda_i t}$, p can only have the form $a + b\,e^{-\mu t}$, with $0 \leqslant a+b \leqslant 1$ and $0 \leqslant a \leqslant 1$. Then

$$\begin{aligned}\frac{dP_n}{dt} &= \binom{N}{n} [np^{n-1}(1-p)^{N-n} - (N-n)\,p^n(1-p)^{N-n-1}]\,(-b\mu\,e^{-\mu t}) \\ &= -\mu b\,e^{-\mu t} \left(\frac{n}{p} - \frac{N-n}{1-p}\right) P_n\end{aligned}$$

and

$$r_n = \frac{N-n}{n+1} \frac{p}{1-p}.$$

Whether or not a solution is possible can be seen by substitution in (4). This produces the sets of equations

$$\alpha_{n+1} \frac{N-n}{n+1} a^2 - (\alpha_n + \beta_n)\,a(1-a) + \beta_{n-1} \frac{n}{N-(n-1)} (1-a)^2 = 0, \tag{5}$$

$$\alpha_{n+1} \frac{N-n}{n+1} 2a - (\alpha_n + \beta_n)\,(1-2a) - \beta_{n-1} \frac{n}{N-(n-1)} 2(1-a) = \mu(Na-n), \tag{6}$$

$$\alpha_{n+1} \frac{N-n}{n+1} + (\alpha_n + \beta_n) + \beta_{n-1} \frac{n}{N-(n-1)} = \mu N. \tag{7}$$

These provide $3(N+1)$ equations for $2(N+1)$ unknowns α_n, β_n, but the sets are not independent. The unique solution satisfying them is the well-known one

$$\alpha_n = \alpha n, \quad \beta_n = \beta(N-n); \qquad a = \beta/(\alpha+\beta), \quad \mu = (\alpha+\beta),$$

and b is determined by the initial conditions.

Polya (1930) considered the compound binomial distribution

$$\begin{aligned}P_n &= \int_0^1 \binom{N}{n} p^n (1-p)^{N-n} \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}\,dp \\ &= \binom{N}{n} \frac{\alpha(\alpha+1)\dots(\alpha+n-1)\,\beta(\beta+1)\dots(\beta+N-n-1)}{(\alpha+\beta)\,(\alpha+\beta+1)\dots(\alpha+\beta+N-1)}.\end{aligned} \tag{8}$$

In his work it arose from a contagion urn scheme in which, after each of N drawings from an urn containing α white balls and β black, two balls of the colour drawn are replaced. P_n is then the probability of obtaining n white balls. Skellam (1948) applies the distribution to the secondary association of chromosomes in *Brassica*, and to accident and traffic problems. It can also be used as a description of the spread of species over N regions when the population size of a species increases by one for each new region in which it is found. The heterogeneity in the binomial then describes that between the abundances of the different species.

The form of P_n, as far as the constants α and β are concerned, is as in a hypergeometric distribution, so that the present result will also apply to that.

The only rational form of α and β that will not make P_n an infinite series in powers of some $e^{-\mu t}$ is $\alpha = a - c\,e^{-kt}$ and $\beta = b + c\,e^{-kt}$, where $0 \leqslant a, b \leqslant 1$, and $c \leqslant a, 1-b$ if $c \geqslant 0$, and $-c \leqslant b, 1-a$ if $c \leqslant 0$. Such an assumption is consistent, for instance, with the spread of a species from a few regions to many, or with the reverse process.

The equations similar to (5), (6) and (7) are now $N+1$ in number, the simplest to find coming from the coefficients of e^{-0kt}, e^{-Nkt} and e^{-kt}. They are:

$$\begin{aligned}\alpha_{n+1}(a+n)(a+n-1)(N-n)(N-(n-1))& \\ -(\alpha_n+\beta_n)(n+1)(N-(n-1))(b+N-(n+1))(a+n-1)& \\ +\beta_{n-1}n(n+1)(b+N-n)(b+N-(n-1)) &= 0, \qquad (9)\end{aligned}$$

$$\begin{aligned}\alpha_{n+1}(N-n)(N-(n-1))+(\alpha_n+\beta_n)(n+1)(N-(n-1))+\beta_{n-1}n(n+1)& \\ = -k(n+1)(N-(n-1)),& \qquad (10)\end{aligned}$$

$$\begin{aligned}&\alpha_{n+1}(N-n)(N-(n-1))(a+n)(a+n-1)[\psi(a+n-1)-\psi(a)-\psi(b+N-(n+1))+\psi(b)] \\ &\quad -(\alpha_n+\beta_n)(n+1)(N-(n-1))(a+n-1)(b+N-(n+1)) \\ &\quad \times[\psi(a+n)-\psi(a)-\psi(b+N-n)+\psi(b)]+\beta_{n-1}n(n+1)(b+N-n)(b+N-(n-1)) \\ &\quad \times[\psi(a+n-1)-\psi(a)-\psi(b+N-(n-1))+\psi(b)] \\ &= -k(n+1)(N-(n-1))(b+N-(n+1))(a+n-1) \\ &\quad \times[\psi(a+n)-\psi(a)-\psi(b+N-n)+\psi(b)], \qquad (11)\end{aligned}$$

where

$$\psi(x) = \frac{d}{dx}\log\Gamma(x).$$

(9) and (10) are satisfied uniquely by

$$\alpha_n = \frac{kn(b+N-n)}{N(a+b+N-1)}, \quad \beta_n = \frac{k(a+n)(N-n)}{N(a+b+N-1)},$$

which do not satisfy (11). Thus this compound binomial cannot arise from a process of this simple birth and death type.

3. THE MORE GENERAL EQUATIONS

When n is not necessarily bounded by N and α_n and β_n are general functions of time, the equations of the process are

$$\left.\begin{aligned}\frac{dP_n}{dt} &= \alpha_{n+1}(t)P_{n+1}-(\alpha_n(t)+\beta_n(t))P_n+\beta_{n-1}(t)P_{n-1}, \\ \frac{dP_0}{dt} &= \alpha_1(t)P_1-\beta_0(t)P_0.\end{aligned}\right\} \qquad (12)$$

The first $n+1$ of these add to give

$$\sum_{r=0}^{n}\frac{dP_r}{dt} = \alpha_{n+1}P_{n+1}-\beta_nP_n \quad (n = 0, 1, \ldots).$$

Thus if it is possible to find a simple closed expression for $\sum_{r=0}^{n}P_r$, we can use this equation to find values for α_n and β_n. Otherwise these values can be found with greater difficulty directly from (12).

3·1. The negative binomial distribution has the form

$$P_n(t) = \frac{\Gamma(k+n)}{\Gamma(k)\,n!}\frac{\sigma^n}{(1+\sigma)^{n+k}} \quad (n = 0, 1, \ldots).$$

Suppose k is constant and the distribution is generated from $P_0(0) = 1$, i.e. $\sigma(0) = 0$. Then, if $\sigma' = d\sigma/dt$,

$$\frac{dP_n}{dt} = \frac{\Gamma(k+n)}{\Gamma(k)\,(n-1)!}\frac{\sigma^{n-1}\sigma'}{(1+\sigma)^{n+k}} - \frac{\Gamma(k+n+1)}{\Gamma(k)\,n!}\frac{\sigma^{n}\sigma'}{(1+\sigma)^{n+k+1}} = f(n-1)-f(n)$$

say, and

$$\sum_0^n \frac{dP_r}{dt} = -P_n\frac{(n+k)\,\sigma'}{1+\sigma} = P_n\left(\alpha_{n+1}\frac{k+n}{n+1}\frac{\sigma}{1+\sigma} - \beta_n\right).$$

The required equation in n is thus, for $n = 0, 1, \ldots,$

$$\frac{-\sigma'}{1+\sigma} = \frac{\alpha_{n+1}}{n+1}\frac{\sigma}{1+\sigma} - \frac{\beta_n}{n+k}. \tag{13}$$

The solution for σ when α_n and β_n are given positive continuous functions is, as required, $\geqslant 0$. The simplest solution is Kendall's (1948*a*), $\alpha_{n+1} = \alpha(n+1)$ and $\beta_n = \beta(n+k)$, where α and β are positive constants. But it is also true that negative binomial will result when α and β vary with time. Generally

$$\sigma = \exp\left(-\int_0^t (\alpha-\beta)\,dt\right)\int_0^t \beta\exp\left(\int_0^t(\alpha-\beta)\,dt\right)dt.$$

Polynomial solutions of higher degrees for α_n and β_n will be more particular solutions in the sense that the coefficients of $n, n^2, \ldots,$ cannot be assigned arbitrarily in both α_n and β_n.

3·2. For the positive binomial with probability parameter p and index N, the equation corresponding to (13) is

$$\frac{dp}{dt} = -\frac{\alpha_{n+1}p}{n+1} + \frac{\beta_n(1-p)}{N-n} \quad (n = 0, 1, \ldots, N-1). \tag{14}$$

A solution $\alpha_{n+1} = \alpha(t)\,(n+1)$, $\beta_n = \beta(t)\,(N-n)$ is then possible for positive functions $\alpha(t)$ and $\beta(t)$. For instance, if n describes the number of regions out of N occupied by a species of plants, it is reasonable to suppose that the rate of new colonization is at its highest after each seeding time, when the rate of loss of colonies is at its lowest, i.e. α_n and β_n are periodic and in anti-phase, a situation typified for example by

$$\left.\begin{aligned}\alpha_n(t) &= (\tfrac{1}{2}\omega)\,(1-\sin\omega t)\,n,\\ \beta_n(t) &= (\tfrac{1}{2}\omega)\,(1+\sin\omega t)\,n.\end{aligned}\right\} \tag{15}$$

For these values, $p = \frac{1}{2}\{1+(1/\sqrt{2})\sin(\omega t - \frac{1}{4}\pi)\}$, having a time lag on $\beta_n(t)$ of half the phase length of α_n and β_n, corresponds to $p(0) = \frac{1}{4}$.

The Poisson equation similar to (13) and (14) is, for a mean m,

$$\frac{dm}{dt} = -\frac{\alpha_{n+1}m}{n+1} + \beta_n,$$

which again is satisfied by arbitrary positive functions $\alpha(t)$ and $\beta(t)$ in $\alpha_n = \alpha(t)\,n$ and $\beta_n = \beta(t)$.

3·3. The geometric distribution is the case $k = 1$ of (13), but the required birth probability rate $\beta_n = \beta(t)\,(n+1)$ is not that usually met in the genesis of a geometric law, *viz.* $\beta_n = \beta(t)\,n$. The difference arises from the different starting points. To get (13) we took $P_0(0) = 1$, and considered a formula for P_n which held for all values of n including 0. Clearly a birth-rate proportional to the population size will not move a process away from $P_0(0) = 1$. If, however, $P_1(0) = 1$, a solution for the birth- and death-rates follows from the supposition that P does not necessarily agree with the formula for P_n with $n = 0$.

Then

$$P_n(t) = \frac{1-P_0(t)}{1+\sigma}\frac{\sigma^{n-1}}{(1+\sigma)^{n-1}} \quad (n = 1, 2, \ldots).$$

The equations obtained from addition of the process equations are

$$\frac{\sigma P_0'}{1-P_0}(1+\sigma) - n\sigma' = \alpha_{n+1}\sigma - \beta_n(1+\sigma) \quad (n = 1, 2, \ldots), \tag{16}$$

and the particular equation

$$\frac{dP_0}{dt} = \alpha_1\frac{1-P_0}{1+\sigma} - \beta_0 P_0. \tag{17}$$

The simplest solutions are those given by Kendall (1948*b*), viz. $\alpha_n = \alpha(t)\,n$ and $\beta_n = \beta(t)\,n$. For these, (16) and (17) reduce to

$$\sigma' = -(\alpha-\beta)\,\sigma + \beta$$

and

$$\frac{P_0'}{1-P_0} = \frac{\alpha}{1+\sigma}, \tag{18}$$

the solutions of which are valid under the initial conditions $P_0(0) = 0$, $\sigma(0) = 0$.

3·4. The logarithmic distribution for the initial conditions $P_1(0) = 1$, $\sigma(0) = 0$, is given by

$$P_n(t) = \frac{1-P_0}{\log(1+\sigma)}\frac{\sigma^n}{n(1+\sigma)^n} \quad (n = 1, 2, \ldots).$$

We require

$$\frac{n\sigma'}{\sigma(1+\sigma)} + \frac{d}{dt}\left[\log\left(\frac{1-P_0}{\log(1+\sigma)}\right)\right] = \frac{\alpha_{n+1}n\sigma}{(n+1)(1+\sigma)} - (\alpha_n+\beta_n) + \frac{\beta_{n-1}n(1+\sigma)}{(n-1)\,\sigma}, \tag{19}$$

together with two special equations for P_1 and P_0. Suppose polynomial solutions are sought for α_n and β_n. Then (19) when multiplied by $n-1$ becomes a polynomial identity in n. Thus α_{n+1} is divisible by $n+1$, and β_{n-1} by $n-1$. Whence, for $n = 0$,

$$1-P_0(t) = \text{constant}\times\log(1+\sigma).$$

This is inconsistent with $P_1(0) = 1$, and $\sigma(0) = 0$; thus polynomial solutions are not possible.

But if $P_0(0) = 1$ and $\sigma(0) = 0$, such solutions exist. An artificial one is the disturbed geometric process, $\alpha_n = \alpha n$, $\beta_n = \beta n$ $(n \geqslant 1)$, where α and β are constants with $\alpha > \beta$; and β_0 is a special immigration term to maintain the population if it should drop to zero, viz.

$$\beta_0 = \frac{\beta k}{1-k\log\left(1+\dfrac{\beta}{\alpha-\beta}(1-e^{-(\alpha-\beta)t})\right)}, \quad \text{with} \quad k < \frac{1}{\log\left(\dfrac{\alpha}{\alpha-\beta}\right)}.$$

Then

$$P_n(t) = \frac{k\sigma^n}{n(1+\sigma)^n} \quad \text{for} \quad n \geqslant 1,$$

$$P_0 = 1-k\log\left[1+\frac{\beta}{\alpha-\beta}(1-e^{-(\alpha-\beta)t})\right], \quad \text{where} \quad \sigma = \frac{\beta}{\alpha-\beta}(1-e^{-(\alpha-\beta)t}).$$

When $k \to 0$, $\beta_0 \to 0$, and $P_0(t) \to 1$, but

$$\frac{P_n}{1-P_0} \to \frac{1}{\log(1+\sigma)}\frac{\sigma^n}{n(1+\sigma)^n} \quad \text{for} \quad n \geqslant 1.$$

4. A STOCHASTIC DEVELOPMENT OF NEW SPECIES

In the previous paragraphs the discussion has been mainly of the growth of a single species. In this paragraph a simple model for the evolution of a number of different species is considered as a possible source of heterogeneity. Yule (1924) discussed the groupings of species into genera when new species arise by mutation. He assumed that the total number of species in a genus is governed by a geometric birth process, so that the probability that a genus contains n species after a time of establishment t is $e^{-\lambda t}(1-e^{-\lambda t})^{n-1}$, for some constant λ. He remarks that one might also suppose that the probability rate of speciation is dependent on the population sizes of the various species, but gives no guide on the form this dependence might take. One might, as a first approximation, suppose that the larger the population of a species, the greater its chance of producing a member which might be classifiable as a member of a new species. We shall assume that each species develops from its original member with birth and death probability rates proportional to the population size n of the species. The extra assumption is that there is a rate of speciation also proportional to this number. Suppose it is kn. There are difficult, incompletely answered questions as to whether there is an exact time of appearance of a new species, what degree of difference it must show from other existing species, and what the effect is of relative geographical isolation of groups of members of a species. We have had to suppose that a species arises at a definite time. Our reply to the second question is that differences of sufficient size are likely to occur when the population size is big enough. Finally, if isolated groups come together again, it seems plausible that the distinctions necessary to stop interbreeding will have most probably occurred at some past time when the total population size was large; so that again the proportionality of speciation rate to size, and the fixing of a definite time of birth of a species are relevant. Then, too, the use of the geometric law for each species enables us, because of its linear birth and death relationship, to consider the population as a whole even though it is split into groups. It is not pretended, however, that the assumptions made are anything but a rough approximation showing, perhaps, the sort of thing that might happen if more factors were taken into account. Thus because of the other factors involved, such as the struggle for a changing amount of food of varying quality, the constant k must be regarded as summing up the mutation rates and the environmental variability.

We shall be concerned only with the number of species and not their genera or higher groupings.

4·1. The distribution of the number of species that have been born in $(0, t)$ is not easily obtainable. Suppose N species have been created at times $\tau_1, \tau_2, \ldots, \tau_N$, and that initially there was one species with one member. Then, if $g_n(t)$ is the probability that a species has n members after time t, the probability of a further birth in $(t, t+dt)$ is

$$k\sum_{i=0}^{N}\sum_{n=0}^{\infty} n(t-\tau_i)\,g_n(t-\tau_i) = k\sum_{i=0}^{N}\overline{n}(t-\tau_i), \quad \text{say.}$$

Suppose

$$k\sum_{i=0}^{j-1}\overline{n}(\tau_j-\tau_i) = \alpha_j(\tau_0, \ldots, \tau_j) \quad \text{with} \quad \tau_0 = 0,$$

and

$$p(0,\tau) = \exp\left(-\int_0^{\tau}\overline{n}(x)\,dx\right).$$

Then it can be shown that $P_{N+1}(t)$, the probability that there are $N+1$ species at time t, is

$$\int_{R_N} p(0,t)\,p(0,t-\tau_1)\dots p(0,t-\tau_N)\prod_{j=1}^{N}\alpha_j(\tau_0,\dots,\tau_j)\,d\tau_1\dots d\tau_N = \int_{R_N} P(\tau_1,\dots,\tau_N,t)\,d\tau_1\dots d\tau_N, \quad \text{say,} \tag{20}$$

where R_N is the region given by the inequalities $0 \leqslant \tau_1 \leqslant \tau_2$, $0 \leqslant \tau_2 \leqslant \tau_3, \dots, 0 \leqslant \tau_N \leqslant t$. $P_{N+1}(t)$ then follows a birth process with transition rate $\beta_{N+1}(t)$ given by

$$\beta_{N+1}(t)\,P_{N+1}(t) = \int_{R_N} P(\tau_1,\dots,\tau_N,t)\,\alpha_{N+1}(\tau_0,\dots,\tau_N,t)\,d\tau_1\dots d\tau_N. \tag{21}$$

It is hardly possible to use these results except for the simplest form of geometric distribution. It would be most satisfactory to deduce a multivariate distribution of the different numbers of species which at time t have $0, 1, \dots$ members. Failing this we can find the average number of species having a given population. Since N follows a birth process, we may write the equation for the mean number, $\overline{N}$, of species at time t,

$$\frac{d\overline{N}}{dt} = \sum_N \beta_N P_N.$$

Then the equation for the expected number N_n of species with n members is

$$\begin{aligned} N_n &= \int_0^t \Sigma\beta_N P_N g_N(t-\tau)\,d\tau + g_n(t) \\ &= \int_0^t \frac{d\overline{N}}{d\tau}\,g_n(t-\tau)\,d\tau + g_n(t). \end{aligned} \tag{22}$$

When this is multiplied by n and summed, our initial assumption gives

$$\frac{d\overline{N}}{dt} = k\left[\int_0^t \frac{d\overline{N}}{d\tau}\,\overline{n}(t-\tau)\,d\tau + \overline{n}(t)\right]. \tag{23}$$

This has the unique solution for the Laplace transforms ϕ of $\overline{n}$, and ψ of $d\overline{N}/dt$

$$\psi = \frac{k\phi}{1-k\phi}. \tag{24}$$

4·2. The simplest example occurs when both birth- and death-rates λ and μ are equal to λ. Then

$$g_n(t) = \frac{(\lambda t)^{n-1}}{(1+\lambda t)^{n+1}} \quad (n = 1, 2, \dots), \qquad g_0(t) = \frac{\lambda t}{1+\lambda t}, \quad \text{with} \quad g_1(0) = 1.$$

This is perhaps the most realistic of the geometric laws, since it corresponds to a group of species so near, on the average, to equilibrium with their environment, that a new species has difficulty in establishing itself. R. A. Fisher (1930) holds that it is reasonable to suppose that most mutant genes that are likely to assist in the development of a species will probably disappear from the population in time. For instance, a gene giving a 1 % advantage will die out with probability 0·98, while one giving no advantage will die out with probability 1. It is commonly considered that most mutants that can survive make little effective change in an organism's relationship to its surroundings. Thus in the present artificial situation in which a new species is supposed to be created at a definite time, it must be assumed that it is not much better adapted than the others. If, as postulated, a species grows geometrically, its chance of possessing no members after time $t \to \mu/\lambda$ if $\lambda > \mu$, and to 1 otherwise. Amongst such laws of growth then, the one most representative of what happens is that for which $\lambda = \mu$. Then $\overline{n}(t) = 1$, and $P_N(t) = e^{-kt}(1-e^{-kt})^{N-1}$ as in

Yule's case, $\overline{N} = e^{kt}$. When the distribution (22) is applied to actual data, the values of kt and λt required are such that the distribution is very near its limiting form

$$\left.\begin{aligned} p_n &= \lim_{t\to\infty} \frac{N_n}{\overline{N}} = \int_0^\infty \frac{\alpha\, e^{-\alpha y}\, y^{n-1}}{(1+y)^{n+1}}\, dy \quad (n = 1, 2, \ldots), \\ p_0 &= 1 - \int_0^\infty \frac{\alpha\, e^{-\alpha y}}{1+y}\, dy, \quad \text{where} \quad \alpha = k/\lambda. \end{aligned}\right\} \tag{25}$$

The p_n depend only on the ratio of k to λ, since these both have the dimension 1/time. The following properties hold:

(i) $p_n > p_{n+1}$ for all $\alpha > 0$, and $n \geqslant 1$.

(ii) p_0 may range from 0 as $\alpha \to \infty$, to 1 as $\alpha \to 0$. As $\alpha \to \infty$, $p_1 \to 1$.

(iii) The rth factorial moment of (25) is $r!(r-1)!/\alpha^{r-1}$. Hence for small k, despite the fact that $p_0(\alpha) \to 1$, the moments become unbounded. This is implied in the increasing average age, $1/\alpha$, of a species, since, as a species grows older, g_n still increases if $n - 1 > 2\lambda \times$ its age, i.e. the larger contributions to the moments increase. When α is large, the ordinary moments all $\to 1$. The overwhelming majority of species are only recently born, and have in virtue of the initial condition for each species only about one member each.

4·3. *Variation in* α. One of the chief defects of the model is the assumption that all species have the same k and λ. In (24) k can be looked on as a mean value, if it is assumed to be independent of time, since it occurs linearly. λ does not occur till (25). In (25) then, it is possible to allow $\alpha = k/\lambda$ to have its own distribution. Because of (i) the resulting distribution of n will have the same general shape as (25).

For (25) there is the recurrence relationship

$$(n+1)\, p_{n+1} = n p_n + \alpha(p_1 + \ldots + p_n) + \alpha^2 e^\alpha \operatorname{Ei}(-\alpha) \quad (n = 1, 2, \ldots), \tag{26}$$

where
$$-\operatorname{Ei}(-\alpha) = \int_\alpha^\infty \frac{e^{-x}}{x}\, dx \quad (\textit{British Association Tables}, \text{vol. 1})$$

and
$$p_1 = \alpha(1 + \alpha\, e^\alpha \operatorname{Ei}(-\alpha)) = \alpha(1 - p_0).$$

The terms extra to np on the right-hand side are of order $\alpha + \alpha^2 e^\alpha \log \alpha$ for small α. The mean value of α will be small, since an ordinary birth is much more common than a species-generating birth. Hence one expects that $E_\alpha p_n$ will vary only a little from a logarithmic form, constant $\times x^n/n$ for small $1-x$. But a small α implies by (iii) a distribution with very large moments. A distribution that leaves all moments of $E_\alpha p_n$ finite does not lead to easily calculable probabilities. The most obvious $dF(\alpha)$ is $dF(\alpha) = (c^m/\Gamma(m))\, e^{-c\alpha} \alpha^{m-1} d\alpha$. Then $[m+1]$ moments are finite. When c is large, as is required by fitted data, the new probabilities may most easily be calculated from

$$\left.\begin{aligned} p_n = \frac{m}{cn} - \frac{m(m+1)}{c^2} \Big[&\log c - \psi(n+1) - \psi(m+2) + \psi(1) + \psi(2) + \ldots \\ &+ \frac{(n+1)\ldots(n+r)\,(m+2)\ldots(m+r+1)}{r!(r+1)!\, c^r} (\log c - \psi(n+r+1) \\ &- \psi(m+r+2) + \psi(r+1) + \psi(r+2)) + \ldots \Big], \\ 1 - p_0 = \frac{m(m+2)}{c(m+1)} &\log c \left(\frac{c}{c-1}\right)^{m+1} + \frac{m(m+2)}{c(m+1)} \\ \times \Big[\psi(1) - \psi(m+1) &+ \ldots + \frac{(m+1)\ldots(m+r)}{r!\, c^r} (\psi(r+1) - \psi(m+r+1)) + \ldots \Big]. \end{aligned}\right\} \tag{27}$$

(27) can be fitted from the first two moments. Thus if u is the observed mean, and v is the ratio of the observed second factorial moment and the mean, the equations are, since the zero class is not observable,

$$\frac{2c}{m-1} = v \tag{28}$$

and

$$\frac{1}{u} = 1-p_0 = -\alpha\, e^{\alpha}\,\mathrm{Ei}(-\alpha). \tag{29}$$

(28) gives c in terms of m; there will only be a solution for m in (29) if

$$\frac{1}{u} > \frac{2}{v}\, e^{2/v}\,(-\mathrm{Ei}(-2/v)).$$

However, in the case tried, namely C. B. Williams's data on different species of Macrolepidoptera caught in light traps over 4 years (Fisher, Corbet & Williams, 1943), a solution is not possible. The nearest approach is when c and $m \to \infty$ in the limiting ratio $2/v$, i.e. when the distribution is back to the form (25). Then a better fit is obtained from the same distribution with the value of α given by the equation $1/u = -\alpha\, e^{\alpha}\,\mathrm{Ei}(-\alpha)$.

Generally, for *any* distribution of α, it is to be expected that the maximum-likelihood fitting of $E_\alpha p_n$ will be very difficult. If, then, one were to fit by moments, including the second, the equation $E_\alpha(2/\alpha) = v$ would be used. Since $E(2/\alpha) \geqslant 2/E(\alpha)$, the smallest mean of $dF(\alpha)$ would be $2/v$, which is reached only when $dF(\alpha)$ tends to the form $P(\alpha \neq \alpha_0) = 0$, $P(\alpha_0) = 1$. For the sort of distribution $dF(\alpha)$ we would be likely to consider, this would mean that the variance of α would $\to 0$ as $E(\alpha) \to 1/E(1/\alpha) = 2/v$. The data, of typical 'log' shape, are fairly well fitted by (25) with $\alpha = 0{\cdot}00290$ and $2/v = 0{\cdot}00326$. Thus it seems likely that a $dF(\alpha)$ leading to a better fit will have to have a mean near $0{\cdot}00290$ and so tending to $0{\cdot}00326$ from above. Thus it will have to have a low variance and there will be little point in using something like (27) rather than (25). An improvement might be found in allowing λ to be different from μ in the geometric law $g_n(t)$.

4·4. *The average distribution with unequal* λ *and* μ. When $\lambda > \mu$ the limiting distribution of the type (22) has infinite moments of all but a few orders, and is thus less satisfactory than (25) as a description of data of this kind.

When $\mu > \lambda$ the distribution is

$$p_n = \frac{(\lambda-\mu+k)\,(\mu-\lambda)}{\mu^2}\left(\frac{\lambda}{\mu}\right)^{n-1}\int_0^1 x^{-1+k/(\mu-\lambda)}\,(1-x)^{n-1}\left(1-\frac{\lambda}{\mu}x\right)^{-(n+1)} dx,$$

in which $k-(\mu-\lambda)$ must be >0, as the average distribution only makes sense if a large number of species has been born ($\bar{N}$ the average number of species born up to time t is proportional to $\exp\,(k-(\mu-\lambda)\,t)$.

If $\xi = k/(\mu-\lambda)$, p_n may be written as

$$\frac{(\xi-1)}{(\mu/(\mu-\lambda))^2}\left(\frac{\lambda}{\mu}\right)^{n-1}\frac{\Gamma(\xi)\,\Gamma(n)}{\Gamma(n+\xi)}\,F(n+1,\xi;\,n+\xi;\,\lambda/\mu) \quad (n = 1, 2, \ldots).$$

If λ/μ is nearly 1, $= \eta/(\eta+1)$ say, the easiest form for calculation for p_0 is, when η is large,

$$\begin{aligned} 1-p_0 &= \frac{\xi-1}{\eta+1}\int_0^1 x^{\xi-1}(1-(\lambda/\mu)\,x)^{-1}\,dx \\ &= \frac{(\xi-1)\,(\eta+1)^{\xi-1}}{\eta^{\xi}}\log\,(\eta+1) - \sum_{r=0}^{\infty}\frac{(\xi-1)\,\xi\ldots(\xi+r-1)}{r!}\,\frac{\psi(\xi+r)-\psi(r+1)}{(\eta+1)^{r+1}}. \end{aligned} \tag{30}$$

Then later p_n's can be calculated from the recurrence relationship

$$(n+1)\,p_n = n\frac{\lambda}{\mu}p_n + \frac{\xi-1}{\eta+1}(p_1+\ldots+p_n) - \frac{\xi-1}{\eta+1}; \quad \left(p_1 = \frac{\xi-1}{\eta+1}p_0\right); \quad (n = 1, 2, \ldots). \tag{31}$$

The chief difference between (31) and (27) is that in (31) the probabilities are exhibited as corrections to a logarithmic term, whereas in (27) they are given as corrections to a harmonic term. Both correction terms are small. The moments of the distribution determined by (31) are all finite. One may find ξ and η from the equations

$$\left.\begin{aligned} \frac{2\eta}{1+\xi} &= v, \\ \xi\int_0^1 \frac{x^{\xi-1}}{1+\eta(1-x)}\,dx &= \frac{1}{u}. \end{aligned}\right\} \tag{32}$$

The values of the right-hand side range from $\log(1+v)/v$ when $\xi = 1$, down to

$$(2/v)\exp(2/v)\,(-\mathrm{Ei}(-2/v))$$

as $\xi \to \infty$ (when η is substituted in terms of ξ). For these data $\xi = 7{\cdot}3$, $\eta = 2549{\cdot}0$. The first 25 frequencies and groupings up to an beyond 50 are compared with those for the log distribution. They agree closely after the first. The figures for the α distribution are so close to those for the ξ, η distribution as not to be worth recording separately. When α is small, as here ($\alpha = 0{\cdot}00290$), it is approximately $1-x$ when the log distribution is given by $N_n = \text{const.}\,x^n/n$. The value of η shows that if this model is at all relevant, the value of λ/μ is almost indistinguishable from 1. The interesting suggestion from the ratio ξ/η is that one species-generating birth occurs about every 350 generations.

Macrolepidoptera figures for all four years

n	N_n obs.	N_n ξ, η distribution	N_n log distribution $x = 0{\cdot}99742$	n	N_n obs.	N_n ξ, η distribution	N_n log distribution $x = 0{\cdot}99742$
1	35	44·09	40·14	17	3	2·25	2·27
2	11	21·80	20·03	18	3	2·11	2·14
3	15	14·38	13·32	19	3	1·98	2·02
4	14	10·54	9·96	20	4	1·87	1·92
5	10	8·36	7·95	21	1	1·77	1·82
6	11	6·90	6·66	22	3	1·68	1·73
7	5	5·87	5·65	23	3	1·60	1·65
8	6	5·09	4·93	24	1	1·52	1·58
9	4	4·49	4·37	25	3	1·45	1·51
10	4	4·01	3·92	26–30	7	6·40	6·71
11	2	3·62	3·56	31–35	3	5·27	5·62
12	2	3·29	3·25	36–40	5	4·46	4·81
13	5	3·02	3·00	41–45	5	3·84	4·19
14	2	2·78	2·77	46–50	3	3·36	3·71
15	4	2·58	2·58	Over 50	60	57·22	63·81
16	3	2·40	2·42	Total	240		

4·5. *Use of the model for a finite distribution.* Suppose as in (1) that species of plants are spreading over M regions. Then the spread may possibly be described by some simple birth and death process (α_n, β_n), with α_n and β_n not depending on time; we might, for instance, consider the logistic process $\alpha_n = \mu n^2$, $\beta_n = \lambda n(M-n)$, where n is the number of regions occupied by a species. If one supposes that the mean number N of species increases exponentially as in 4·2 and 4·4, the average rate of production of new species is

$$\sum_N \beta_N(t)\, P_N(t) = a\, e^{at},$$

where a is a constant. Then the age distribution of the $\overline{N} = e^{at}$ species $\rightarrow a\, e^{-ax}\, dx$ as $t \rightarrow \infty$. Suppose the probability that a species of age x occupies n regions is $g_n(x)$ when $g_1(0) = 1$. The limiting distribution p_n that we have been using can then be obtained directly from the equations of the process for the $g_n(x)$. Thus

$$\begin{aligned} a\int_0^\infty e^{-at}\frac{dP_n(t)}{dt}\,dt &= a^2\int_0^\infty e^{-at}\,P_n(t)\,dt - a_n P_n(0) \\ &= \alpha_{n+1}a\int_0^\infty e^{-at}\,P_{n+1}\,dt - (\alpha_n+\beta_n)\,a\int_0^\infty e^{-\alpha t}\,P_n\,dt \\ &\qquad + \beta_{n-1}a\int_0^\infty e^{-at}\,P_{n-1}\,dt \quad (N \geqslant n > 1), \end{aligned} \tag{33}$$

or, as

$$a\int_0^\infty e^{-at}\,P_n\,dt = p_n,$$

$$ap_n = \alpha_{n+1}p_{n+1} - (\alpha_n+\beta_n)\,p_n + \beta_{n-1}p_{n-1}, \tag{34}$$

with the special end equations,

$$\begin{aligned} -a + ap_1 &= \alpha_2 p_2 - (\alpha_1+\beta_1)\,p_1 + \beta_0 p_0, \\ ap_0 &= \alpha_1 p_1 - \beta_0 p_0, \end{aligned}$$

and

$$ap_N = -\alpha_N p_N + \beta_{N-1}p_{N-1}.$$

The equations add to give a recurrence relationship such as (26) or (31),

$$\alpha_{n+1}p_{n+1} = \beta_n p_n + a(p_0 + \ldots + p_n) - a \quad (n = 1, 2, \ldots). \tag{35}$$

When p_0 can be found, as is the case in 4·2, (35) can be used to give the p_n's. For $n \leqslant M$, p_n is of the form $A_n p_0 + B_n$, the coefficients A_n and B_n being obtained from (35). Then p_n can be found by addition of the probabilities.

To show the possibility of such 'explanation' of the distribution of species, we can fit data (kindly supplied by Prof. A. R. Gemmell) on the spread of moss species over the six islands of Hawaii. The process $\alpha_n = 0$, $\beta_n = \lambda n(M-n)$ has been used. This is not quite unrealistic, as a species is only observed when it has a large number of plants and so is not so likely to disappear. However, even with $\alpha_n = 0$, the fits are good:

$$\left.\begin{aligned} p_n &= \frac{(n-1)\,(M-(n-1))}{\gamma+n(M-n)}\,\frac{(n-2)\,(M-(n-2))}{\gamma+(n-1)\,(M-(n-1))}\cdots\frac{1(M-1)}{\gamma+2(M-2)}\,\frac{\gamma}{\gamma+1(M-1)} \\ p_1 &= \frac{\gamma}{\gamma+1(M-1)}, \quad \gamma = k/\lambda. \end{aligned}\right\} \tag{36}$$

The figures are, for the three different reproductive types of species, dioecious, monoecious and sterile:

Moss species in Hawaii

Dioecious			Monoecious			Sterile		
n	Obs.	Exp.	n	Obs.	Exp.	n	Obs.	Exp.
1	12	11·19	1	15	13·88	1	22	21·78
2	6	5·70	2	6	5·64	2	7	7·19
3	4	4·22	3	3	3·39	3	5	3·56
4	3	3·87	4	2	2·48	4	1	2·11
5	4	4·53	5	0	2·13	5	0	1·39
6	13	12·49	6	4	2·47	6	2	0·97
$\gamma = 1{\cdot}816$			$\gamma = 4{\cdot}308$			$\gamma = 7{\cdot}156$		

γ was calculated in each case from the maximum-likelihood equation

$$\sum_n N_n\left(\frac{1}{\gamma+n(M-n)}+\dots+\frac{1}{\gamma+1(M-1)}\right) = (\sum_n N_n)\frac{1}{\gamma}.$$

4·6. *Remarks.* The simple assumption of proportionality of speciation rate to the total population thus leads to distributions substantially like those observed. Yule's data for the increasing number of species in a genus bears out the exponential manner of increase found for N. For such an N the form of the p_n can always be found as in 4·5, and more realistic growth laws for the different species may account for any divergences in observed data from a logarithmic form. For instance, random immigration of new species or of members of existing species (which might include back mutation to other already existing species) may be considered. But these particular effects are of small weight in the p_n unless their rate of occurrence is of as high an order as the rate of increase of N. A more important change might be the consideration of competitive populations described by something like the logistic law, but they are possibly best considered in relation to the shifting of the population inside the living region.

I wish to thank Prof. M. S. Bartlett for the suggestions he made in the course of this work, and Mr A. M. Walker for his helpful comments.

REFERENCES

FISHER, R. A. (1930). *Genetical Theory of Natural Selection.* Oxford University Press.
FISHER, R. A., CORBET, A. S. & WILLIAMS, C. B. (1943). *J. Anim. Ecol.* **12**, 42.
KENDALL, D. G. (1948*a*). *Biometrika*, **35**, 6.
KENDALL, D. G. (1948*b*). *Ann. Math. Statist.* **19**, 1.
POLYA, G. (1930). *Ann. Inst. Poincaré*, **1**, 117.
SKELLAM, J. G. (1948). *J.R. Statist. Soc.* B, **10**, 257.
WILLIAMS, C. B. (1944). *J. Ecol.* **32**, 1.
YULE, G. U. (1924). *Phil. Trans.* B, **213**, 21.

MODIFICATIONS TO THE VARIATE-DIFFERENCE METHOD

By M. H. QUENOUILLE

Institute of Statistics, University of Oxford

1. Introduction

1·1. Methods of trend elimination in time-series analysis fall into two main groups. First, where the form of trend is known or may be assumed to be polynomial, it is possible to fit polynomials and to use residuals from this fitting in further analysis. This method has been described elsewhere (Quenouille, 1952).

A second, more exploratory, method which assumes nothing about the form of the trend makes use of moving average difference formulae for its estimation. This method is, however, subject to the criticisms that the form and extent of moving averages chosen would seem to be somewhat arbitrary, and that the elimination process may easily give rise to misleading results (Spencer-Smith, 1947; Quenouille, 1949).

1·2. The variate-difference method (Anderson, 1914) is closely related to the method of moving averages, and may be regarded as having greater generality. Any moving average difference formula may be shown to be equivalent to an average of higher order differences and, consequently, results obtainable from such formulae may be obtained by appropriate manipulation of the variate-differences. For example, if $\nabla^p x_i$ is used to denote the pth variate difference, i.e.

$$\nabla^p x_i = x_{i+p} - \binom{p}{1} x_{i+p-1} + \binom{p}{2} x_{i+p-2} - \dots,$$

then a simple five-term moving average difference formula may be written

$$\begin{aligned} 5d_5 &= -x_{i+2} - x_{i+1} + 4x_i - x_{i-1} - x_{i-2} \\ &= -\nabla^2 x_i - 3\nabla^2 x_{i-1} - \nabla^2 x_{i-2}. \end{aligned}$$

Results obtainable using the residuals from such a formula are therefore derivable from the variate-differences.

1·3. The classical variate-difference method has therefore a greater flexibility than the use of moving averages, since the effect of the latter may be judged from the variate-difference. Further, it has the advantage that criteria and tests of significance can be set up to judge the extent of the trend.

A major disadvantage still exists in that the existence of serial correlation in the residual element of the series will bias the results and produce misleading conclusions about the form of the series. Also, it is not clear that the tests used when serial correlation is not present in the residual element are the best possible.

1·4. The following paper sets out to do two things. The first part, §§2–6, is concerned with deriving more sensitive tests of the existence of trend in variate-differences of any order, and with finding more accurate estimates of variances and covariances of the residual element from any set of variate-differences assumed to be trend free. The latter part, §§7–10, is concerned with how this theory should be modified if serial correlation is assumed to exist in the residuals. As such it does not give a strict criterion for the separation of trend and serial correlation, but it does provide a basis on which the separation of the two may take place.

For example, the analysis of § 7·3 indicates the existence of both trend and serial correlation in Kendall's sheep series. It shows further that trend is unlikely to be important in the third and higher variate-differences. The analysis of § 9·5 shows further that, *if* it is accepted that there is no trend in the third and higher variate-differences, at least a second-order autoregressive process is needed to represent the series and that, *if* such is used, the variance of the trend in the second variate-differences is unlikely to be more than 0·4. The conclusion is therefore that a second-order autoregressive scheme with a trend, small in its second differences, is a good representation of the series. As emphasized elsewhere (Quenouille, 1949), this is not, however, the only possible representation, and it would still be possible to maintain that the series consisted of a complicated trend with a random element or, possibly, of a high-order autoregressive scheme with no trend. The process to be outlined gives rise to the simplest possible representation for any time series.

2. General investigation

2·1. The classical variate-difference method is based upon the idea that by taking successive differences of a series of observations, the 'smooth' part or trend of the observations will be eliminated, and what remains can be used to find out about the random element. In particular, if the series $x_1, x_2, \ldots, x_{N+p}$ is free from serial correlation and no trend component exists in the pth variate-difference, $\nabla^p x_i$, then

$$\frac{(\nabla^p x_i)^2}{\binom{2p}{p}}$$

estimates the variance, σ^2, of the random element.

Any average of a series of such terms also estimates σ^2, and, commonly, estimates such as

$$V_p = \frac{\sum_{i=1}^{N} (\nabla^p x_i)^2}{N\binom{2p}{p}}$$

are used for this purpose.

For reasons to be seen later, we shall use a modified definition which applies *end-corrections* to the averaging process.

This will not affect the large sample theory, but it will have the joint effects of making it more applicable to small samples and of simplifying conversion formulae.

2·2. We shall define a series of summation functions $\Sigma_{(i)}$ by the formulae:

$$\begin{aligned}
\Sigma_{(0)} x_i y_i &= x_1 y_1 + x_2 y_2 + \ldots + x_n y_n, \\
\Sigma_{(1)} x_i y_i &= \tfrac{1}{2} x_1 y_1 + x_2 y_2 + \ldots + x_{n-1} y_{n-1} + \tfrac{1}{2} x_n y_n, \\
\Sigma_{(2)} x_i y_i &= \tfrac{1}{4} x_1 y_1 + \tfrac{3}{4} x_2 y_2 + \ldots + \tfrac{3}{4} x_{n-1} y_{n-1} + \tfrac{1}{4} x_n y_n,
\end{aligned}$$

the summation in each case covering all values up to x_n, y_n. The general law of formation is determined by the recurrence relation

$$\sum_{i=1}^{n}{}_{(m)} x_i y_i = \sum_{i=1}^{n-1}{}_{(m-1)} \tfrac{1}{2}(x_i y_i + x_{i+1} y_{i+1}),$$

so that, for example, the first three coefficients in $\Sigma_{(3)}$ are $\frac{1}{8}$, $\frac{1}{2}$ and $\frac{7}{8}$.

2·3. For any series, $x_1, x_2, \ldots, x_m$, the estimated variances, ${}_mV_0, {}_mV_1, {}_mV_2, \ldots$, will be defined by

$$\begin{aligned} {}_mV_0 &= \frac{\Sigma_{(m)} x^2}{N}, \\ {}_mV_1 &= \frac{\Sigma_{(m-1)} (\nabla x)^2}{2N}, \\ \vdots \quad & \qquad \vdots \\ {}_mV_p &= \frac{\Sigma_{(m-p)} (\nabla^p x)^2}{N\dbinom{2p}{p}}. \end{aligned}$$

Under this definition, p cannot exceed m. However, m may be arbitrarily chosen so that it can always be arranged that it is equal to or greater than the largest value of p required.

In the following sections, m will be taken as equal to the largest value of p occurring. This is usually ten. Further, the prefix m will, in general, be omitted unless the line of argument requires its inclusion.

2·4. Serial correlation coefficients may be similarly defined, so that

$${}_mr_p = \frac{\Sigma_{(m-p)} x_i x_{i+p}}{\Sigma_{(m)} x_i^2}.$$

Again, m must be chosen to equal or exceed p, and, with this condition, the suffix m may usually be omitted.

2·5. It is then not very difficult to derive the following identities (see Kendall, 1946, p. 23):

$$V_p = V_0\left[1 - \frac{2p}{p+1} r_1 + \frac{2p(p-1)}{(p+1)(p+2)} r_2 - \frac{2p(p-1)(p-2)}{(p+1)(p+2)(p+3)} r_3 - \ldots\right]$$

and

$$r_p V_0 = V_0 - p^2 V_1 + \frac{p^2(p^2-1)}{(2!)^2} V_2 - \frac{p^2(p^2-1)(p^2-4)}{(3!)^2} V_3 + \ldots.$$

These formulae demonstrate the basic equivalence of the approaches employing variate-differences and serial correlations. They also provide a simple way in which to find the variances of combinations of the V_i under different conditions. For instance,

$$\operatorname{var} V_p = \frac{2\sigma^4}{N}\left[1 + \frac{2p^2}{(p+1)^2} + \frac{2p^2(p-1)^2}{(p+1)^2(p+2)^2} + \ldots\right]$$

for a random series.

2·6. When a set of variate-differences have been calculated as described above, it is reasonable to ask what is the best estimate of the residual or error variance in any series that may be obtained using these estimates. If it is supposed that $V_{m+1} \ldots V_{m+p}$ are free from any trend component, then the problem is to find coefficients c_i $(i = m+1 \ldots m+p)$ so that $\sum_{i=m+1}^{m+p} c_i = 1$ and so that $\sum_{i=m+1}^{m+p} c_i V_i$ has the smallest possible variance.

This problem may be solved by finding a restricted minimum. If we note that

$$\frac{N}{\sigma^4}\left[\operatorname{cov}(V_i, V_j) - \frac{\kappa_4}{N}\right] = \frac{2\dbinom{2i+2j}{i+j}}{\dbinom{2i}{i}\dbinom{2j}{j}} = a_{ij}, \text{ say}$$

where κ_4 is the fourth cumulant of the residual elements, and will normally be assumed to be zero, and $(\kappa_4+\lambda\sigma^4)/N$ is supposed to be the variance of $\sum_{i=m+1}^{m+p} c_i V_i$, the coefficients c_i are given by the symmetrical equations

$$\left.\begin{aligned} \sum_{i=m+1}^{m+p} a_{ji}c_i - \lambda &= 0 \\ -\sum_{i=m+1}^{m+p} c_i \qquad &= -1 \end{aligned}\right\} \quad (j = m+1, \ldots, m+p).$$

Solution of these equations gives rise to the values for c_i and λ.

2·7. The following particular solutions exist for these equations

m	$p=1$		$p=2$		
	c_{m+1}	λ	c_{m+1}	c_{m+2}	λ
0	1	3·00	2·50	−1·50	2·50
1	1	3·89	3·86	−2·86	3·00
2	1	4·62	5·20	−4·20	3·43
3	1	5·25	6·54	−5·54	3·82
4	1	5·82	7·88	−6·88	4·17
5	1	6·33	9·21	−8·21	4·49
6	1	6·81	10·55	−9·55	4·79

m	$p=3$				$p=4$				
	c_{m+1}	c_{m+2}	c_{m+3}	λ	c_{m+1}	c_{m+2}	c_{m+3}	c_{m+4}	λ
0	4·67	− 7·00	3·33	2·33	7·50	− 20·25	22·50	− 8·75	2·25
1	10·36	− 17·91	8·55	2·69	22·71	− 67·28	72·25	− 26·68	2·52
2	18·18	− 33·04	15·86	3·00	50·37	− 152·56	161·08	− 57·90	2·77
3	28·14	− 52·43	25·29	3·28	94·33	− 287·71	300·79	−106·42	3·00
4	40·23	− 76·09	36·86	3·55	159·00	− 486·23	505·16	−176·92	3·21
5	54·45	−104·01	50·56	3·79	245·2	− 749·6	774·4	−269·0	3·41
6	70·79	−136·17	66·38	4·02	356·6	−1088·9	1119·9	−386·7	3·60

It may be seen that fairly substantial improvements in accuracy can be achieved by using several values of V_i simultaneously. For instance, the variance of V_4 is 5·25. By combining V_4 and V_5 the variance is reduced to 3·82, while the use of V_4 to V_7 reduces it to 3·00.

2·8. It should be noted that the coefficients, c_i, do not differ appreciably from those that would be used if the a_{ij} were regarded as functions of i for which the value corresponding to $i=-1$ was required. For instance, in this case, for $p=2$, we should have $c_{m+1}=m+3$, $c_{m+2}=-(m+2)$; for $p=3$,

$$c_{m+1} = (m+3)(m+4)/2, \quad c_{m+2} = -(m+2)(m+4), \quad c_{m+3} = (m+2)(m+3)/2.$$

This provides rough empirical approximations to the coefficients, c_i, but for practical purposes it is convenient to deal with a still rougher approximation, derived by extrapolating the value of V_i corresponding to $i=0$, using the formula

$$V_0 = V_m - m\nabla V_m + \frac{m(m+1)}{2!}\nabla^2 V_m - \ldots.$$

If this is used, the following coefficients are obtained:

m	$p = 1$		$p = 2$			$p = 3$				$p = 4$				
	c_{m+1}	λ	c_{m+1}	c_{m+2}	λ	c_{m+1}	c_{m+2}	c_{m+3}	λ	c_{m+1}	c_{m+2}	c_{m+3}	c_{m+4}	λ
0	1	3·00	2	−1	2·56	3	−3	1	2·42	4	−6	4	−1	2·35
1	1	3·89	3	−2	3·08	5	−6	2	2·89	8	−15	11	−3	2·75
2	1	4·62	4	−3	3·53	10	−15	6	3·17	20	−45	36	−10	2·98
3	1	5·25	5	−4	3·93	15	−24	10	3·48	35	−84	70	−20	3·24
4	1	5·82	6	−5	4·29	21	−35	15	3·76	56	−140	120	−35	3·48
5	1	6·33	7	−6	4·62	28	−48	21	4·03	84	−216	189	−56	3·71
6	1	6·81	8	−7	4·93	36	−63	28	4·28	120	−315	280	−84	3·92

2·9. This table shows that an appreciable increase in the efficiency of estimation of the residual variance is obtained even with simple extrapolation. The efficiency of this extrapolation method as compared with the most efficient linear combinations, obtained in § 2·7, is as follows:

m	$p = 1$	$p = 2$	$p = 3$	$p = 4$
0	100 %	98 %	96 %	96 %
1	100 %	97 %	93 %	92 %
2	100 %	97 %	95 %	93 %
3	100 %	97 %	94 %	93 %
4	100 %	97 %	94 %	92 %
5	100 %	97 %	94 %	92 %
6	100 %	97 %	94 %	92 %

The loss in efficiency increases with both m and p. However, in view of the computational convenience and other properties, to be given later, of the extrapolation method, its use is generally to be advocated. Any loss in efficiency from its use can be compensated by employing more values of V_i.

2·10. If it may be presumed that $V_{m+1} \ldots V_{m+p}$ are free from trend, the trend component in V_m may be estimated by comparing it with $V_{m+1} \ldots V_{m+p}$. A further question is then posed: which estimate of the trend component in V_m is the most accurate? To answer this, we have to find coefficients $d_i\,(i = m+1, \ldots, m+p)$, subject to $\sum\limits_{i=m+1}^{m+p} d_i = 1$, such that the variance, $(\kappa_4 + \lambda\sigma^4)/N$, of $V_m - \sum\limits_{i=m+1}^{m+p} d_i V_i$, is a minimum. These coefficients may be shown to satisfy the equations

$$\left.\begin{aligned} \sum_{i=m+1}^{m+p} (a_{ji} - a_{mi})\, d_i - \lambda &= a_{jm} - a_{mm} \\ -\sum_{i=m+1}^{m+p} d_i \qquad\qquad &= -1. \end{aligned}\right\} \quad (j = m+1, \ldots, m+p)$$

Solution of these equations gives rise to values for d_i and λ.

2·11. The following particular solutions exist for these equations:

m	$p = 1$		$p = 2$			$p = 3$			
	d_{m+1}	λ	d_{m+1}	d_{m+2}	λ	d_m	d_{m+1}	d_{m+2}	λ
1	1	0·222222	2·327	−1·327	0·030612	4·081	−5·388	2·306	0·007752
2	1	0·108889	2·230	−1·230	0·006949	3·763	−4·636	1·873	0·000924
3	1	0·067347	2·177	−1·177	0·002469	3·587	−4·239	1·652	0·000202
4	1	0·046838	2·143	−1·143	0·001113	3·475	−3·992	1·517	0·000063
5	1	0·034974	2·121	−1·121	0·000582	3·402	−3·834	1·432	0·000022
6	1	0·027391	2·104	−1·104	0·000337	3·345	−3·713	1·368	0·000010

It may be seen that these coefficients tend to binomial coefficients, i.e. $d_{m+i} \doteqdot (-1)^{i-1}\binom{p}{i}$, and this may be proved mathematically without much difficulty. (In the limit, the coefficients $a_{ji} - a_{mi}$ may be accurately represented as polynomials in j. It follows that a limiting solution is given by the differencing process, $d_{m+i} = (-1)^{i-1}\binom{p}{i}$ and $\lambda = 0$.) Estimates of trend components are thus given by variate-differences of the V_i.

2·12. The function $\sum_{i=m}^{m+p} (-1)^{m+i}\binom{p}{i-m} V_i$ thus estimates the trend in V_m, if $V_{m+1} \ldots V_{m+p}$ are free from trend. Its variance is given by

$$\frac{2\sigma^4}{N} \sum_{i=0}^{p} \sum_{j=0}^{p} (-1)^{i+j} \frac{\binom{p}{i}\binom{p}{j}\binom{4m+2i+2j}{2m+i+j}}{\binom{2m+2i}{m+i}\binom{2m+2j}{m+j}} = \frac{\lambda\sigma^4}{N}.$$

This is of order $m^{\frac{1}{2}-2p}$. The following table gives the values of λ:

m	Values of p					
	0	1	2	3	4	5
1	3·0000	0·2222	0·042222	0·015918	0·007927	0·004602
2	3·8889	0·1089	0·010522	0·002394	0·000796	0·000330
3	4·6200	0·0673	0·003935	0·000599	0·000142	0·000043
4	5·2531	0·0468	0·001833	0·000200	0·000035	0·000007
5	5·8187	0·0350	0·000980	0·000080	0·000011	0·000002
6	6·3346	0·0274	0·000577	0·000037	0·000004	—
7	6·8118	0·0222	0·000364	0·000019	—	—
8	7·2578	0·0185	0·000242	—	—	—
9	7·6781	0·0157	—	—	—	—
10	8·0766	—	—	—	—	—

For completeness, the variances of the V_i have been included in this table in the column $p = 0$.

2·13. The efficiency of estimation of trend components using this approach is given in the following table:

m	1	2	3
1	100 %	73 %	49 %
2	100 %	66 %	39 %
3	100 %	63 %	34 %
4	100 %	61 %	31 %
5	100 %	59 %	28 %
6	100 %	58 %	27 %

It is again true that the efficiency decreases as m and p are increased. However, the loss in efficiency may be easily compensated by including one or two extra values of V_i, i.e. by increasing p.

2·14. The two processes of estimating the residual variance and the trend components may thus be carried out using differences of the quantities V_i. However, it must be remarked that this is not the only possible approximate basis for estimation. An alternative, which is likely to be somewhat better, would acknowledge the fact that the variance of V_m increases roughly linearly with $\sqrt{(m)}$ by extrapolating to zero using a suitably chosen scale instead of m.

One such scale appears to be provided by $\sqrt{(m+1)}-1$. If this is employed, with the method of divided differences, more efficient estimates can be obtained. The extra computation would, however, appear to be seldom worth while.

2·15. It might be asked what is the probable bias resulting from the failure to eliminate trend completely.

Suppose, for instance, that extrapolation is carried out using $V_m \dots V_{m+p}$, but that a trend component σ_t^2 is present in V_m. The error in the extrapolated estimate, V_0', is then

$$\sigma_t^2+m\sigma_t^2+\dots+\frac{m(m+1)\dots(m+p-1)}{p!}\sigma_t^2=\frac{(m+1)(m+2)\dots(m+p)}{p!}\sigma_t^2.$$

However, the variance of σ_t^2 is

$$\frac{2\sigma^4}{N}\sum_{i=0}^{p}\sum_{j=0}^{p}(-1)^{i+j}\frac{\binom{p}{i}\binom{p}{j}\binom{4m+2i+2j}{2m+i+j}}{\binom{2m+2i}{m+i}\binom{2m+2j}{m+j}}=\frac{\lambda\sigma^4}{N}.$$

Thus the variance of the error in V_0' is

$$\left[\frac{(m+1)(m+2)\dots(m+p)}{p!}\right]^2\frac{\lambda\sigma^4}{N}=\frac{\lambda'\sigma^4}{N},\quad\text{say.}$$

λ' is of order $\sqrt{(m)}/p$.

The following table gives a series of values for λ':

m	p				
	1	2	3	4	5
1	0·89	0·38	0·25	0·20	0·17
2	0·98	0·38	0·24	0·18	0·15
3	1·08	0·39	0·24	0·17	0·13
4	1·17	0·41	0·24	0·17	0·11
5	1·26	0·43	0·25	0·17	0·13
6	1·34	0·45	0·26	0·18	—
7	1·42	0·47	0·27	—	—
8	1·50	0·49	—	—	—
9	1·57	—	—	—	—

As a consequence of the fact that λ' decreases as p increases, we may say that the bias in the estimate of V_0 decreases with $\sqrt{(p)}$.

3. EXAMPLES

3·1. This theory may be illustrated using the two series given in the following table. The figures are in hundreds:

i	1	2	i	1	2	i	1	2	i	1	2
1	1000	1018	14	273	467	27	74	471	40	20	170
2	905	927	15	247	465	28	67	458	41	18	148
3	819	847	16	223	465	29	61	442	42	17	128
4	741	776	17	202	468	30	55	423	43	15	109
5	670	714	18	183	473	31	50	402	44	14	93
6	607	661	19	165	477	32	45	378	45	12	78
7	549	615	20	150	483	33	41	353	46	11	65
8	497	576	21	135	487	34	37	327	47	10	54
9	449	543	22	122	490	35	33	299	48	9	44
10	407	518	23	111	492	36	30	272	49	8	36
11	368	498	24	100	491	37	27	245	50	7	29
12	333	483	25	91	488	38	25	219	51	7	25
13	301	472	26	82	481	39	22	193	—	—	—

Both of these series consist of a trend plus a random element.

The trend in series 1 is given by

$$T_i = 10^5 \exp\left(-\frac{i-1}{10}\right) \quad (i = 1, \ldots, 51),$$

and that in series 2 by

$$T_i = 10^5\left[\exp\left(-\frac{i-1}{10}\right) + \frac{1}{\sqrt{(2\pi)}}\exp -\frac{(i-26)^2}{200}\right] \quad (i = 1, \ldots, 51).$$

In both series the random element consists of the error committed in reading the functions from three-place tables. Thus, in series 1,

$$u_i = 10^5\left[\exp\left(-\frac{i-1}{10}\right) \text{ to three decimal places } - \exp\left(-\frac{i-1}{10}\right)\right],$$

and, in series 2,

$$u_i = 10^5\Big[\exp\left(-\frac{i-1}{10}\right) \text{ to three decimal places} - \exp\left(-\frac{i-1}{10}\right) + \frac{1}{\sqrt{(2\pi)}}\exp-\frac{(i-26)^2}{200} \text{ to three decimal places } - \frac{1}{\sqrt{(2\pi)}}\exp-\frac{(i-26)^2}{200}\Big].$$

The value of σ^2 is consequently 816·6 for series 1 and 1633·3 for series 2.

Table 3·1*a*. *Estimates of the trend components in series* 1

V		Estimates				
		1st	2nd	3rd	4th	5th
1	2591687*	2582023·2	2573539·8	2565029·0	2556536·6	2548067·4
2 2	(± 699) 9663·8* (± 795)	8483·4* (± 42·1)	8510·8* (± 13·08)	8492·4	8468·2	8442·6
3	1180·4 (± 867)	− 27·4 (± 33·1)	18·4 (± 8·00)	24·2* (± 3·12)	25·6* (± 1·52)	26·0* (± 0·84)
4	1207·8 (± 924)	− 45·8 (± 27·6)	− 5·8 (± 5·46)	− 1·4 (± 1·80)	− 0·4 (± 0·75)	− 0·1 (± 0·34)
5	1253·6 (± 973)	− 40·0 (± 23·9)	− 4·4 (± 3·99	− 1·0 (± 1·14)	− 0·3 (± 0·42)	− 0·3 (± 0·18)
6	1293·6 (± 1015)	− 35·6 (± 21·1)	− 3·4 (± 3·06)	− 0·7 (± 0·78)	0·0 (± 0·26)	— —
7	1329·2 (± 1053)	− 32·2 (± 19·0)	− 2·7 (± 2·43)	− 0·7 (± 0·56)	— —	— —
8	1361·4 (± 1087)	− 29·5 (± 17·3)	− 2·0 (± 1·98)	— —	— —	— —
9	1390·9 (± 1118)	− 27·5 (± 16·0)	— —	— —	— —	— —
10	1418·4 (± 1146)	— —	— —	— —	— —	— —

* Significant at upper 1 % level.

Table 3·1*b*. *Estimates of the trend components in series* 2

V		Estimates				
		1st	2nd	3rd	4th	5th
1	2173658·54*	2154432·32	2136689·82	2119064·59	2101548·55	2084135·64
2	19226·22*	17742·50*	17625·23	17516·04	17412·91	17311·91
3	1483·72	117·27	109·19*	103·13*	101·00*	100·23*
4	1366·45	8·08	6·06	2·13	0·77	0·36
5	1358·37	2·02	3·93	1·36	0·41	0·15
6	1356·35	− 1·91	2·57	0·95	0·26	—
7	1358·26	− 4·48	1·62	0·69	—	—
8	1362·74	− 6·10	0·93	—	—	—
9	1368·84	− 7·03	—	—	—	—
10	1375·87	—	—	—	—	—

3·2. Tables 3·1*a* and 3·1*b* give estimates of $V_1 - V_{10}$ for the two series together with the values of successive differences $\nabla^m V_i$ ($m = 1$ to 5).

The bracketed figures in table 3·1*a* give the standard errors of the estimates, calculated using $\sigma^2 = 816{\cdot}7$. Those for table 3·1*b* would be $\sqrt{(2)} = 1{\cdot}4$ times these values.

3·3. Examination of the estimates for series 1 shows a trend component of about 26 in V_3 and a negligible component in V_4. Similarly, for series 2, a trend component of about 100 exists in V_3 and a negligible component in V_4.

Estimators of V_0 may therefore be formed using values after V_3. Some possible combinations are given in the following table:

	Series 1	Series 2
$V_4 - V_9$	919·0	1549·1
$V_5 - V_{10}$	893·8	1522·6
$V_6 - V_{10}$	969·4	1484·9
$V_7 - V_{10}$	969·4	1421·9
$V_8 - V_{10}$	1053·4	1349·9

The coefficients of variation for those estimates vary between roughly

$$100\sqrt{(2{\cdot}8141)} = 26{\cdot}1\ \%$$

for the first pair and

$$100\sqrt{(4{\cdot}5141)} = 33{\cdot}1\ \%$$

for the last pair. None of these estimates differs from its theoretical value by more than its standard error.

3·4. To set more exact confidence limits, it is necessary to calculate the efficiency factors, E, of the estimators, using the values of λ given in the table of section 2·8. These are given by $E = 2/\lambda$.

Confidence limits may then be calculated in the same manner as for an ordinary estimate of variance based upon EN observations.

For instance, the efficiency value for the fourth estimate is $2/3{\cdot}92 = 0{\cdot}5102$, so that the effective number of observations is $41 \times 0{\cdot}5102 = 21$. The 95 % limits for σ^2 for series 1 are thus $969{\cdot}4/1{\cdot}69 = 573{\cdot}6$ and $969{\cdot}4 \times 2{\cdot}04 = 1977{\cdot}6$.

4. CORRELATION ANALYSIS FOR RANDOM SERIES

4·1. For any two series it is possible to calculate not only the V's estimating the residual variances of the series, but also a set of C's estimating the residual covariances between the series. These C's may be calculated in an exactly similar manner to the V's except that each square has to be replaced by a product.

For example, if two series x_i and y_i ($i = 1, \ldots, N+10$) are used, it is necessary to replace $(\nabla^m x_i)^2$ by $(\nabla^m x_i)(\nabla^m y_i)$ in calculating C_m. Thus, for instance, we may set up a series of equations

$$C_{10} = \frac{(\nabla^{10}x_1)(\nabla^{10}y_1) + (\nabla^{10}x_2)(\nabla^{10}y_2) + \ldots + (\nabla^{10}x_N)(\nabla^{10}y_N)}{N},$$

$$C_9 = \frac{\frac{1}{2}(\nabla^9 x_1)(\nabla^9 y_1) + (\nabla^9 x_2)(\nabla^9 y_2) + \ldots + \frac{1}{2}(\nabla^9 x_{N+1})(\nabla^9 y_{N+1})}{N},$$

etc.

4·2. A curious analogue exists between the formulae for the covariances and those for the variances. The pth covariance may be written in the form

$$C_p = \mathrm{cov}_0(x, y) - \frac{2p}{p+1}\mathrm{cov}_1(x, y) + \frac{2p(p-1)}{(p+1)(p+2)}\mathrm{cov}_2(x, y) - \ldots,$$

where

$$\mathrm{cov}_0(x, y) = \mathrm{cov}(x_i y_i),$$

$$\mathrm{cov}_1(x, y) = \tfrac{1}{2}[\mathrm{cov}(x_i y_{i+1}) + \mathrm{cov}(x_{i+1}, y_i)],$$

$$\mathrm{cov}_2(x, y) = \tfrac{1}{2}[\mathrm{cov}(x_i y_{i+2}) + \mathrm{cov}(x_{i+2}, y_i)].$$

The formula for V_p is the particular case of this where $x_i = y_i$.

4·3. Now, if the correlation between x and y is ρ, and no serial or lagged correlation exists in the two series, it is not difficult to show that

$$\mathrm{var}[\mathrm{cov}_0(x, y)] = (1+\rho^2)\frac{\sigma_x^2\sigma_y^2}{N}$$

$$\mathrm{var}[\mathrm{cov}_i(x, y)] = \frac{1+\rho^2}{2}\frac{\sigma_x^2\sigma_y^2}{N} \quad (i = 1, 2, \ldots).$$

$$\mathrm{cov}[\mathrm{cov}_i(x, y), \mathrm{cov}_j(x, y)] = 0.$$

Since the factor $(1+\rho^2)$ occurs in all of these expressions, any process which minimizes the variance of a combination of the C_i's does so *irrespective of the value of* ρ. Consequently, the processes derived for the V_i (i.e. where $\rho = 1$) may also be applied to the C_i; the residual covariance between x and y may be estimated by extrapolation in the same manner as the residual variances were estimated.

4·4. For the same reason, correlation coefficients may be calculated using variances and covariances which have been estimated in similar fashions. The variances of these will be

$$\frac{(1-\rho^2)^2}{EN},$$

where E represents the efficiency of estimation.

The values of this factor are given in the following table, assuming that extrapolations involving V_{m+1} to V_{m+p} or C_{m+1} to C_{m+p} are used in estimating the correlation coefficient. It is derived by dividing the values of λ in the table of §2·8 into two (the minimum possible value for λ).

Efficiency of estimated correlation coefficient

m	1	2	3	4
0	0·667	0·781	0·826	0·851
1	0·514	0·649	0·692	0·727
2	0·433	0·567	0·631	0·671
3	0·381	0·509	0·575	0·617
4	0·344	0·466	0·532	0·575
5	0·316	0·433	0·496	0·539
6	0·294	0·406	0·467	0·510

4·5. It seems unlikely that the distribution of correlation coefficients calculated by the extrapolation method would differ very greatly from the distribution of ordinary correlation coefficients, since, in the limit, the estimates are 100 % efficient. It would therefore seem to be reasonable to test any correlation calculated by the extrapolation method as if it were an ordinary correlation coefficient based upon En pairs of observations. This should give a good approximation to the true state of affairs.

5. EXAMPLE

5·1. An example may be provided using series 1 and 2 of the last section. These series have a theoretical covariance of 816·7 and a theoretical correlation of 0·707.

Table 5·1*a* gives estimates of C_1 to C_{10} and their differences.

Table 5·1*a*. *Estimates of the covariances between series* 1 *and series* 2

p	C_p	Differences				
		1st	2nd	3rd	4th	5th
1	1561690·09	−1551285·16	1542029·34	−1532770·58	1523540·43	−1514343·00
2	10410·93	−9255·82	9258·76	−9230·15	9197·43	−9163·66
3	1155·11	2·94	28·61	32·72	33·77	−34·15
4	1158·05	31·55	−4·11	1·05	−0·38	0·19
5	1189·60	27·44	−3·06	0·67	−0·19	0·06
6	1217·04	24·38	−2·39	0·48	−0·13	—
7	1241·42	21·99	−1·91	0·35	—	—
8	1263·41	20·08	−1·56	—	—	—
9	1283·49	18·52	—	—	—	—
10	1302·01	—	—	—	—	—

This table shows similar characteristics to tables 3·1*a* and 3·1*b*, trend being largely eliminated in fourth and higher differences.

5·2. Estimates of covariances may be obtained in the same manner as estimates of variances. For example, using C_4–C_9, we get

$$\begin{aligned} C_0 &= C_4 - 4\nabla C_4 + \frac{4\times 5}{2}\nabla^2 C_4 - \frac{4\times 5\times 6}{6}\nabla^3 C_4 + \ldots \\ &= 1158{\cdot}05 - 4\times 31{\cdot}55 - 10\times 4{\cdot}11 - 20\times 1{\cdot}04 - 35\times 0{\cdot}38 - 56\times 0{\cdot}19 \\ &= 945{\cdot}8. \end{aligned}$$

The following table gives a group of estimates obtained in this fashion:

Estimate using	Estimated covariance
C_4–C_9	945·8
C_5–C_{10}	962·2
C_6–C_{10}	977·3
C_7–C_{10}	1004·6
C_8–C_{10}	1046·6

5·3. These estimates may be used to find the correlation between the two series. The following table gives a number of estimates of correlation coefficients together with the effective number of pairs of observations on which they are based.

Estimate using	Coefficients	No. of pairs
C_1*	0·658	27
C_2*	0·764	21
C_3*	0·873	18
C_4	0·901	16
C_5	0·912	14
C_6	0·919	13
C_7	0·924	12
C_8	0·928	11
C_9	0·930	11
C_{10}	0·932	10
C_4–C_9	0·793	29
C_5–C_{10}	0·811	27
C_6–C_{10}	0·815	23
C_7–C_{10}	0·856	21
C_8–C_{10}	0·878	18

* Biased by the existence of trend components.

It may be seen that, apart from the first three estimates, which are biased by the trend, the estimates obtained using serial estimates of covariance are better than those individually obtained.

5·4. Approximate limits may be set for any one coefficient by the use of the z transformation. For instance, the value of z corresponding to the coefficient 0·793 is 1·080. This has a standard error of roughly $1/\sqrt{(29)} = 0{\cdot}186$, and its 95 % confidence limits are consequently roughly 0·715 and 1·445. The corresponding 95 % confidence limits for ρ are 0·41 and 0·89.

It is perhaps worth noting that the actual sample correlation between the two series when their trend is removed is 0·781. The agreement with this calculated correlation coefficient is very good.

6. THE EFFECT OF SERIAL CORRELATION

6·1. The theory given above is based upon the assumption that the series used are serially independent. Where this assumption does not hold, the accuracy of the estimates of variance, covariance and correlation will differ from the values given above. For this reason, it is necessary to consider whether serial correlation exists in the series used and, if so, what modifications in the ordinary procedures need to be adopted.

In this and the next sections, attention will be largely concentrated upon the methods of answering three questions. These are:

(1) How may serial correlations in the series be detected?

(2) May any existing serial correlations be fitted into a scheme of the moving average or autoregressive type?

(3) What procedures exist for testing the correlation between series where both trend and serial correlation exist?

6·2. Before starting on a more serious consideration of these questions, one important point might be noted: If a large number of variate-differences are used in estimating the variances or covariances, these will give in general unbiased estimates of variances and covariances.

This will be apparent if it is remembered that

$$V_p = V_0\left[1-\frac{2p}{p+1}r_1+\frac{2p(p-1)}{(p+1)(p+2)}r_2-\dots\right],$$

and that the estimate of V_0 obtained from V_{m+1} to V_{m+p}, by either the most exact or by the extrapolation method, will be of the form

$$V_0' = V_0[1-a_1r_1-a_2r_2-a_3r_3-\dots],$$

where $a_i \to 0$ as $p \to \infty$.

A more rigorous proof is outlined in Appendix 1. It indicates a desirable asymptotic independence of the serial correlation in the series.

6·3. Of course, in practice, we cannot let p become very large, but, even so, an appreciable reduction in bias may be achieved by the use of several of the V's (or C's).

For example, if $V_1, V_2, \dots, V_p$, are used, the most exact method gives an estimate

$$V_0' = V_0\left[1-\frac{r_1}{p}-\frac{r_2}{p}-\dots-\frac{r_p}{p}\right],$$

while the extrapolation method gives an estimate

$$V_0' = V_0\left[1-\frac{2}{p+1}r_1-\frac{4(p-1)}{(p+1)(p+2)}r_2-\frac{6(p-1)(p-2)}{(p+1)(p+2)(p+3)}r_3-\dots\right].$$

Both of these are liable to differ very little from V_0.

6·4. To illustrate the actual sizes of the bias, the following table gives the values that would be obtained for σ^2 by the extrapolation methods on different series for which $\sigma^2 = 1$:

Estimate using	Series			Moving average	Moving average
	Markoff	Auto-regressive	Auto-regressive		
	$\rho = 0{\cdot}5$	$a = 1{\cdot}1$, $b = 0{\cdot}5$	$a = 1{\cdot}1$, $b = 0{\cdot}8$	$\rho_1 = 0{\cdot}5$	$\rho_1 = -0{\cdot}64$, $\rho_2 = 0{\cdot}30$
V_1	0·500	0·267	0·389	0·500	1·640
V_5	0·365	0·067	0·039	0·167	2·352
V_{10}	0·349	0·058	0·031	0·091	2·648
V_1–V_{10}	0·796	0·841	1·179	0·909	1·035
V_2–V_{10}	0·668	0·594	1·071	0·818	1·096
V_5–V_{10}	0·477	0·175	0·201	0·545	1·445
V_8–V_{10}	0·386	0·078	0·047	0·273	2·040

These values are reasonably satisfactory for V_1–V_{10} and V_2–V_{10}, but are poor for V_8–V_{10}.

6·5. The first of the above queries may be reformulated to read: If $V_{m+1} \dots V_{m+p}$ may be assumed to be without any trend component and $\rho_{q+1}, \rho_{q+2} \dots$ may be assumed zero, how may we estimate $\rho_1, \rho_2 \dots \rho_q$?

Let us consider, first of all, the case $q = 1$.

One approach would be to express the means and variances of $V_{m+1} \dots V_{m+p}$ in terms of ρ_1 and σ_0, and thence to obtain maximum likelihood, or for rough uses least square, estimates of ρ_1 and σ. These would be quite complicated to obtain and would still provide only a first step towards simplifying the serial correlation picture; it would undoubtedly be necessary to reconsider these estimates as ρ_2 and higher serial correlations were taken into account. A simplification of the procedure, to allow more rapid approximate testing to be carried out, is therefore desirable.

One possible simplification is to ignore in the first instance the dependence of the variances of the V's on the ρ_i. It would then be possible to estimate $\sigma^2, \rho_1\sigma^2, \dots$ by least square procedure.

An equivalent method to this is to estimate $\rho_1 V_0$ from the differences

$$V_{m+1} - V_{m+2}, \quad V_{m+2} - V_{m+3}, \dots, V_{m+p-1} - V_{m+p},$$

and to use this to correct for the bias in the value estimated for V_0 on the assumption that $\rho_1 = 0$. This approach is the simplest mathematically and numerically.

6·6. We now consider a series of functions, $R_{11}, R_{12}, \dots$ defined by the formula

$$R_{1p} = \frac{(p+1)(p+2)}{1 \cdot 2} \nabla V_p.$$

It is not hard to show that

$$R_{1p} = V_0 \left[r_1 - \frac{4p}{p+3} r_2 + \frac{9p(p-1)}{(p+3)(p+4)} r_3 - \frac{16p(p-1)(p-2)}{(p+3)(p+4)(p+5)} r_4 + \dots \right],$$

so that if $\rho_2 = \rho_3 = \dots = 0$, we have

$$E(R_{1p}) = \rho_1 \sigma^2,$$

$$\frac{N}{\sigma^4} \operatorname{cov}(R_{1p} - R_{1q}) = \frac{1}{2} \frac{(2p+2q)!}{(p+q+2)!\,(p+q)!} \frac{(p+2)!\,p!}{2p!} \frac{(q+2)!\,q!}{2q!} \frac{3pq+2p+2q+1}{4pq+2p+2q+1}.$$

6·7. Using this latter value, it is possible to set up equations similar to those of § 2·6. These may then be used to give the coefficients c_i and λ for estimating the first serial covariance and its variance. The following table gives the values of λ:

m	p			
	1	2	3	4
0	2·00	1·50	1·33	1·25
1	3·92	2·37	1·88	1·64
2	6·73	3·54	2·58	2·11
3	10·54	5·03	3·43	2·52
4	15·42	6·86	4·44	2·91
5	21·47	9·05	5·65	4·54

6·8. These values of λ might be compared with those obtained by extrapolation:

m	p		
	2	3	4
0	1·52	1·39	1·32
1	2·56	2·15	1·85
2	4·15	3·29	2·88
3	6·37	4·88	4·03
4	9·28	6·93	5·38
5	12·95	9·56	8·71

In this instance, direct extrapolation is not as efficient as previously, owing to the rapid increase in the variance of R_{1p} with p. This is of order $(p+1)^{\frac{5}{2}}$.

Better estimates may be obtained using the device suggested in §2·14 of using an alternative scale as a basis for extrapolation. In this instance, something like a quadratic scale is suggested, and the use of the scale $(m+3)^2$ appears to be fairly efficient.

6·9. It is fairly apparent that anything like a full analysis to estimate the higher serial correlation coefficients would involve a great deal of work which could not in general be justified. The prime need is for some fairly rapid method of gauging the likely serial correlations in any series using their variate differences. The above analysis gives an indication of how this may be approached.

In general, it is possible to define a series of functions

$$R_{mp} = \frac{(p+2m-1)(p+2m)}{(2m-1)\,2m} \nabla R_{m-1.p},$$

such that if the serial correlations higher than the mth are zero, the expectation of R_{mp} is $\rho_m \sigma^2$. These may be used for general investigational purposes.

The variances of the R_{mp} increase vary rapidly with p, so that the higher values are unreliable and, for extrapolation purposes, cubic, quartic and higher-order scales are required.

7. Examples

7·1. As a first illustration of these functions, series 1 and 2 may be used. Tables 7·1*a* and 7·1*b* give the estimates, R_{mp}, of serial correlation for these two series. Examination of these

Table 7·1*a*. *Estimates of serial covariance for series* 1

	0	1st	2nd	3rd	4th	5th
1	2591687	7746069·6	12825282·0	17776284·4	22551645·6	27102772·4
2	9663·8	50900·4	127936·0	236115·6	376650·0	545402·0
3	1180·4	− 274·0	1445·5	1755·6	4785·0	6921·2
4	1207·8	− 687·0	714·0	− 680·4	792·0	− 723·3
5	1253·6	− 840·0	940·8	−1016·4	1149·7	−5888·9
6	1293·6	− 996·8	1218·0	−1429·1	3673·5	—
7	1329·2	−1159·2	1542·8	−2559·4	—	—
8	1361·4	−1327·5	2035·0	—	—	—
9	1390·9	−1512·5	—	—	—	—
10	1418·4	—	—	—	—	—

Table 7·1*b*. *Estimates of serial covariance for series* 2

	0	1st	2nd	3rd	4th	5th
1	2173658·54	6463296·96	10594736·62	14464143·22	17973894·37	21032116·20
2	19226·22	106455·00	263205·75	484447·60	765799·30	1098472·34
3	1483·72	1172·70	3680·25	7950·26	16840·88	28245·34
4	1366·45	121·20	367·64	− 623·28	545·49	233·08
5	1358·37	42·42	575·40	− 854·70	430·23	935·46
6	1356·35	− 53·48	808·50	−1009·14	29·32	—
7	1358·26	−161·28	1037·85	−1018·16	—	—
8	1362·74	−274·50	1233·65	—	—	—
9	1368·84	−386·65	—	—	—	—
10	1375·87	—	—	—	—	—

tables is sufficient to show that, in both instances, trend apparently exists in the third variate-difference and that the first three serial correlation coefficients are likely to be negligible. At first sight, the largest would appear to be the first serial correlation coefficient of series 2, but an extrapolation using R_{15}–R_{19} gives an estimate of 117·32 for the serial covariance. This is well within the bounds of chance variation.

7·2. This analysis was carried out for four other theoretical series. These were:

Series 3. Eighty terms of a Markoff process with $\rho = 0·5$. In fact, the first eighty terms of Kendall's (1949) series 11.

Series 4. Eighty terms of a moving average process for which $\rho_1 = 0·4$. This was determined by $u_i = 2\epsilon_i + \epsilon_{i-1}$, where ϵ_i was the random element from Kendall's series 11.

Series 5. Eighty terms of the autoregressive process $u_i = 1·1u_{i-1} - 0·8u_{i-2} + \epsilon_i$. In fact, the first eighty terms of Kendall's (1945) series 3 multiplied by 0·05.

Series 6. Series 5 plus a trend component determined from

$$T_i = \exp -\frac{i}{50} + \sin 5i^0 + \frac{(i-30)^2}{1000}.$$

The estimates of serial correlation for these series are given in Tables 7·2*a*–7·2*d*.

Table 7·2*a*. *Estimates of serial covariance for series* 3 ($\sigma^2 = 1·33$)

p \ m	0	1st	2nd	3rd	4th	5th
1	0·812338	0·404673	0·267895	0·055272	0·035729	0·149848
2	0·677447	0·243936	0·228415	0·027483	−0·086874	0·075819
3	0·636791	0·152570	0·213692	0·081538	−0·138569	0·019698
4	0·621534	0·091515	0·179718	0·152082	−0·149928	0·109313
5	0·615433	0·053004	0·129024	0·215688	−0·203984	0·290341
6	0·612909	0·031500	0·070200	0·288913	−0·328416	—
7	0·611784	0·022140	0·004538	0·389964	—	—
8	0·611169	0·021645	−0·070455	—	—	—
9	0·610688	0·028050	—	—	—	—
10	0·610178	—	—	—	—	—

Table 7·2*b*. *Estimates of serial covariance for series* 4 ($\sigma^2 = 5\cdot0$)

ρ \ m	0	1st	2nd	3rd	4th	5th
1	3·561117	2·503398	0·273930	−0·306278	−0·239615	0·615038
2	2·726651	2·339040	0·492700	−0·119911	−0·742828	0·328054
3	2·336811	2·141960	0·556938	0·342293	−0·966501	0·200368
4	2·122615	1·982835	0·414316	0·834330	−1·082098	0·452106
5	1·990426	1·894053	0·136206	1·293402	−1·305667	1·215830
6	1·900233	1·871352	−0·216540	1·762103	−1·826737	—
7	1·833399	1·900224	−0·617018	2·324176	—	—
8	1·780615	1·967535	−1·063975	—	—	—
9	1·736892	2·064260	—	—	—	—
10	1·699360	—	—	—	—	—

Table 7·2*c*. *Estimates of serial covariance for series* 5 ($\sigma^2 = 0\cdot0905$)

ρ \ m	0	1st	2nd	3rd	4th	5th
1	0·02762389	0·04650171	0·03204585	0·00373254	−0·02157966	−0·02510146
2	0·01212332	0·02727420	0·02937975	0·02051672	−0·00104210	−0·01603533
3	0·00757762	0·01552230	0·01838865	0·02116514	0·00989108	−0·00416622
4	0·00602539	0·01026840	0·00956984	0·01612968	0·01229467	0·00632286
5	0·00534083	0·00821772	0·00419328	0·01091376	0·00916798	0·00983670
6	0·00494951	0·00751884	0·00121680	0·00762269	0·00495225	—
7	0·00468098	0·00735660	−0·00051563	0·00609892	—	—
8	0·00447663	0·00741285	−0·00168850	—	—	—
9	0·00431190	0·00756635	—	—	—	—
10	0·00417433	—	—	—	—	—

Table 7·2*d*. *Estimates of serial covariance for series* 6 ($\sigma^2 = 0\cdot0905$)

ρ \ m	0	1st	2nd	3rd	4th	5th
1	0·02943038	0·05195976	0·04107950	0·01625029	−0·00570086	−0·00601408
2	0·01211046	0·02731206	0·02947215	0·02068429	−0·00078025	−0·01561504
3	0·00755845	0·01552320	0·01839128	0·02116978	0·00986637	−0·00416615
4	0·00600613	0·01026855	0·00957054	0·01614690	0·01226992	0·00656084
5	0·00532156	0·00821772	0·00418824	0·01094148	0·00902555	0·00874398
6	0·00493024	0·00751968	0·00120420	0·00770154	0·00527813	—
7	0·00466168	0·00735912	−0·00054615	0·00607750	—	—
8	0·00445726	0·00741870	−0·00171490	—	—	—
9	0·00429240	0·00757460	—	—	—	—
10	0·00415468	—	—	—	—	—

Examination of these tables shows that in no case could we be in doubt about the existence of serial correlation. There is very little difference between Tables 7·2*c* and 7·2*d*, and, from these tables alone, there might easily be some doubt about the existence of a trend component in series 6.

Table 7·2*c* indicates what is likely to be a common characteristic of oscillatory autoregressive schemes: the direct estimates of serial covariances are greater in many cases than the corresponding estimates of variance. This is due to the fact that the latter will tend to be greatly reduced by the serial correlation, while the former will not be reduced to the same extent and may even be increased.

7·3. As practical illustrations of this approach, it is convenient to consider Kendall's (1943) ungraduated sheep series for 1867–1939 and Davis's (1941) figures of monthly freight-car loadings for the United States.

Estimates of serial covariances for the former series are given in Table 7·3*a*.

Table 7·3*a*. *Estimates of serial covariance for Kendall's sheep series*

ρ \ m	0	1st	2nd	3rd	4th	5th
1	3177·8585	6065·7162	5180·0270	2607·2578	875·0051	1306·9226
2	1155·9531	2957·7000	3317·7000	1926·6983	−194·2952	−946·0034
3	663·0031	1630·6200	2285·5402	2047·5931	450·7071	−424·8454
4	499·9411	977·6085	1432·3764	1818·1422	695·8102	−1471·9133
5	434·7672	670·7607	826·3290	1522·9500	1423·6794	−2034·3657
6	402·8305	532·9492	410·9790	1011·8856	2295·5504	—
7	383·7966	478·1520	181·0050	305·5624	—	—
8	370·5146	458·4060	122·2430	—	—	—
9	360·3278	447·2930	—	—	—	—
10	352·1952	—	—	—	—	—

There would seem to be little doubt about the existence of either a trend component or serial correlation in this series.

Trend apparently may be regarded as being confined to the first two variate differences, in which case an estimate of σ^2 may be derived from the third and higher variate differences. This estimate is more than 3700.

There is a marked similarity between the general characteristics of these estimates and those of series 5 and 6. The large values taken by R_{1p} and R_{2p} indicate that a second-order autoregressive scheme is likely to be needed to represent this series.

7·4. The analysis of Davis's freight-car loadings requires special adjustment because of the existence of seasonal variation. This could be easily carried out by an extension of the analysis of variance technique. It required the extraction of seasonal components from the formulae of § 2·3, using $m = 11$ instead of 10 as usual. Starting with 167 observations, this gave us a value for N, $167-11-12 = 144$.

Table 7·4*a*. *Estimates for serial covariance for freight-car loadings*

p \ m	0	1st	2nd	3rd	4th	5th
1	900·290	835·608	415·660	265·937	320·355	483·472
2	621·754	586·212	225·705	16·772	−75·213	62·040
3	524·052	495·930	216·720	63·571	−117·513	68·763
4	474·459	434·010	190·232	123·396	−157·184	−65·781
5	445·525	393·246	149·100	190·080	−124·655	−108·848
6	426·799	368·396	97·260	234·828	−78·006	—
7	413·642	355·428	43·890	258·830	—	—
8	403·769	350·640	−5·885	—	—	—
9	395·977	351·175	—	—	—	—
10	389·592	—	—	—	—	—

The estimates of serial covariance for this series, when seasonal variation has been eliminated, are given in Table 7·4*a*.

There is clearly a high trend component in V_1, but apparently not much beyond this. There is also clearly a high first serial correlation coefficient, a low second serial correlation coefficient and not much beyond this. These values suggest the fitting of either a Markoff process or a moving average scheme of length two. There is a fair degree of resemblance between this table and Table 7·2*a*.

8. CORRELATION ANALYSIS FOR NON-RANDOM SERIES

8·1. Before discussing correlation analysis for non-random series, it might be noted that the serial covariances of the variate differences may be directly calculated using the V_i. If D_{pq} is used to denote the qth serial covariance for the pth variate difference, then

$$D_{pq} = \frac{\Sigma_{(m-p-q)}(\nabla^p x_i)(\nabla^p x_{i+q})}{N\binom{2p}{p}},$$

and the following identities apply:

$$D_{p1} = V_p - \frac{2p+1}{p+1} V_{p+1},$$

$$D_{p2} = V_p - 4\frac{2p+1}{p+1} V_{p+1} + 2\frac{(2p+1)(2p+3)}{(p+1)(p+2)} V_{p+2},$$

$$D_{p3} = V_p - 9\frac{2p+1}{p+1} V_{p+1} + 12\frac{(2p+1)(2p+3)}{(p+1)(p+2)} V_{p+2} - \frac{4(2p+1)(2p+3)(2p+5)}{(p+1)(p+2)(p+3)} V_{p+3}.$$

These may be used to calculate the serial correlations of the variate differences up to any order.

8·2. A direct application of these formulae occurs when we consider the effect on the V_i of any linear transformation of the x_i. For instance, if the transformation

$$y_i = (x_i - \rho x_{i-1})/\sqrt{(1+\rho^2)}$$

is carried out, the revised V_i are given by

$$V_i' = V_i - \frac{2\rho}{1+\rho^2} D_i.$$

Similarly, a transformation of the type

$$y_i = (x_i - a x_{i-1} + b x_{i-2})/\sqrt{(1+a^2+b^2)}$$

gives values

$$V_i' = V_i - \frac{2a(1+b)}{1+a^2+b^2} D_{i1} + \frac{2b}{1+a^2+b^2} D_{i2}.$$

8·3. Correlation analysis may be carried out for non-random series in a similar fashion to that for random series. However, unless at least one of the series is random, the significance of the resulting coefficients is liable to be incorrectly estimated.

As an alternative approach to testing the correlation between non-random series, it is therefore often convenient to try to transform the series so that they consist of uncorrelated random elements. This is the approach which will be considered.

The major alternative to this approach is to adjust the degrees of freedom of the resulting estimates by a factor which would take into account the serial correlation in the series. This might be done using Bartlett's (1935) adjustment. (See, for example, Quenouille (1952).)

8·4. What the best method may be for determining a suitable transformation for serially correlated values when trend exists is not clear. However, one approach which appears to be reasonably satisfactory is to employ the transformation so that estimates of serial covariance of §6 disappear. This may be done using either the individual values or the extrapolates. The method will be clarified by examples.

9. EXAMPLES

9·1. It is not difficult to calculate the effect on Table 7·2*a* of adding D_{p1} to the estimates, V_p. This effect is given in Table 9·1*a*. The estimate of the first serial covariance from the first four values of Table 7·2*a* is 0·673841, while the corresponding estimate of the increment from Table 9·1*a* is 0·831860. Thus, 0·673841/0·831860 = 0·81 times Table 9·1*a* if subtracted from Table 7·2*a* gives a zero estimate for the first serial correlation coefficient.

Table 9·1*a*

p \ m	0	1st	2nd	3rd	4th	5th
1	−0·203832	0·540117	0·229965	0·151753	0·099939	0·054527
2	−0·383872	0·402138	0·121570	0·074023	0·055326	−0·019595
3	−0·450894	0·353510	0·082915	0·039598	0·056662	−0·086132
4	−0·486245	0·329820	0·066416	0·010752	0·106354	—
5	−0·508233	0·315588	0·062832	−0·034368	—	—
6	−0·523261	0·305116	0·075720	—	—	—
7	−0·534158	0·295020	—	—	—	—
8	−0·542353	—	—	—	—	—

This corresponds to a transformation $y_i = (x_i - \rho x_{i-1})/(1+\rho^2)$, where $2\rho/(1+\rho^2) = 0{\cdot}81$, i.e. $\rho = 0{\cdot}51$, and, if this is made in the above instance, all of the serial correlations may obviously be neglected.

9·2. The existence of trend in the above series would have necessitated the use of higher terms. For instance, if trend existed in V_1, the appropriate factor might have been estimated using the second to fifth terms. The estimate would have been 0·55982/0·685038 = 0·817, i.e. $\rho = 0{\cdot}52$.

9·3. For more complicated schemes, it is necessary to calculate the perturbations due to $D_{p1} D_{p2}$ and to consider these jointly. These may be used to set up simultaneous equations to find coefficients necessary to make the higher order serial covariances equal to zero.

For instance, for series 6, the perturbations due to D_{p1} and D_{p2} are as in Table 9·3*a*. Since more than the values in table 9·3*a* would need to be subtracted from Table 7·2*d* to obtain zero estimates, it is apparent that a Markoff scheme would not suffice as a means of representing the series and that at least a second autogressive scheme must be used.

Table 9·3*a*

Perturbation due to D_{p1}

p \ m	0	1st	2nd	3rd	4th	5th
1	0·01126469	0·03525495	0·03410505	0·01768935	0·00511876	0·00283112
2	−0·00048696	0·01479192	0·02146980	0·01370809	0·00280239	−0·00792403
3	−0·00295228	0·00620400	0·01412618	0·01196438	0·00820514	−0·00277375
4	−0·00357268	0·00216795	0·00914102	0·00778722	0·00980538	—
5	−0·00371721	0·00020916	0·00654528	0·00362736	—	—
6	−0·00372717	−0·00088172	0·00555600	—	—	—
7	−0·00369568	−0·00162252	—	—	—	—
8	−0·00365061	—	—	—	—	—

Perturbation due to D_{p2}

p \ m	0	1st	2nd	3rd	4th	5th
1	−0·00544013	−0·00659004	0·01186480	0·02297309	0·02621533	0·02378685
2	−0·00324345	−0·01370892	−0·00454455	0·00258399	0·00675336	0·01197115
3	−0·00095863	−0·01189110	−0·00592883	−0·00161810	−0·00140879	0·00229327
4	0·00023048	−0·01019715	−0·00525462	−0·00090090	−0·00273183	—
5	0·00091029	−0·00907116	−0·00495432	0·00035806	—	—
6	0·00134225	−0·00824544	−0·00502470	—	—	—
7	0·00163673	−0·00757548	—	—	—	—
8	0·00184716	—	—	—	—	—

To estimate the appropriate proportions of Tables 9·3*a* and 9·3*b* to be added to Table 7·2*d*, extrapolates based upon the third to sixth variate differences might be used. These give equations

$$0{\cdot}06902037+0{\cdot}04286921A-0{\cdot}02305761B = 0,$$

$$0{\cdot}07588594+0{\cdot}05124778A-0{\cdot}01022722B = 0,$$

or

$$A = -1{\cdot}40,$$

$$B = 0{\cdot}38.$$

These values might be compared with the true values $A = -1{\cdot}39$, $B = 0{\cdot}56$ and with those obtained using extrapolates based upon the second to sixth variate differences, $A = -1{\cdot}39$, $B = 0{\cdot}11$.

9·4. The insensitivity of B in this instance is of no great concern so far as subsequent correlation analyses is concerned, since it is apparent that any inaccuracies in A and B will give rise to schemes for which the serial correlation coefficients are small. It is, however, preferable to choose values of A and B corresponding to schemes for which the coefficients a and b are real so that subsequent analysis may be carried out if required. This necessitates that both $1+A+B$ and $8B(1-B)-A$ should exceed zero. Thus, for instance, no scheme exists for which $A = -1{\cdot}40$, $B = 0{\cdot}38$, but, if $A = -1{\cdot}40$ and $B = 0{\cdot}44$, then both conditions are satisfied and $a = 1{\cdot}6$, $b = 1{\cdot}0$ gives a possible scheme.

The adaption of this latter scheme instead of the true scheme means that the quantities used in any correlation analysis are still serially correlated. The amount of this serial correlation is, however, too small ($\rho_1 = 0{\cdot}06$, $\rho_2 = -0{\cdot}29$, $\rho_3 = -0{\cdot}27$, $\rho_4 = -0{\cdot}07$) to be of any importance.

The variate difference analysis may then be derived by the appropriate multiples of the above table from Table 7·2*d*. This gives the following results:

m p	0	1st	2nd	3rd	4th	5th
1	0·01126616	−0·00029679	−0·00144706	0·00159336	−0·00133238	0·00048857
2	0·01136509	0·00057145	−0·00258517	0·00262992	−0·00173212	0·00074591
3	0·01126984	0·00160552	−0·00399406	0·00370768	−0·00224069	0·00072614
4	0·01110929	0·00274667	−0·00553892	0·00484840	−0·00265962	—
5	0·01092618	0·00393359	−0·00715505	0·00597672	—	—
6	0·01073887	0·00512609	−0·00878507	—	—	—
7	0·01055579	0·00629744	—	—	—	—
8	0·01038086	—	—	—	—	—

This table indicates that very little serial correlation now exists in the series. The differences of the revised values of V_i may then be used to gain a better idea of the trend in the new series.

9·5. A similar analysis was carried out using Kendall's sheep series. This indicated clearly that a Markoff process could not be used to represent the series. The equations for determining A and B, derived using the third to sixth variate differences were

$$7437{\cdot}9107 + 4292{\cdot}2967A - 2408{\cdot}4016B = 0,$$

$$6891{\cdot}9200 + 5904{\cdot}1121A + 2639{\cdot}7281B = 0,$$

from which $A = -1{\cdot}42$, $B = 0{\cdot}56$.

The corresponding estimates using the second to sixth variate differences were $A = -1{\cdot}35$, $B = 0{\cdot}78$.

Both of these estimates are markedly greater than those for the artificial series, though the former estimate is very near the expected values for the artificial series. This serves to show that the differencing process $u_n - 1{\cdot}1u_{n-1} + 0{\cdot}8u_{n-2}$ will give rise to nearly random components which may then be used in a correlation analysis. The following table gives the values of V_i if $A = -1{\cdot}4$, $B = 0{\cdot}5$, i.e. if $a = 1{\cdot}21$, $b = 0{\cdot}76$. It may be seen that with these values there is very little evidence of serial correlation and trend does not seem to be of any importance after the first variate difference:

m p	0	1st	2nd	3rd	4th	5th
1	934·9359	72·5463	122·1597	211·8888	95·6965	457·7057
2	910·7538	−0·7502	− 29·1894	137·4583	−278·7898	658·7784
3	910·8788	10·9256	−102·8280	310·9275	−727·9566	−4415·6187
4	909·7863	40·3050	−232·3812	249·9381	1819·5162	—
5	907·0993	90·1909	−315·6938	−90·3925	—	—
6	902·8087	142·7166	−291·0414	—	—	—
7	897·7117	181·5221	—	—	—	—
8	892·6694	—	—	—	—	—

This latter statement can, of course, be tested using differences of the first column. In this instance there appear to be components of the order of 25·5 and 0·4 in V_1 and V_2.

9·6. A final analysis was carried out using Davis's freight-car loadings. Here again, the serial correlations could not be accounted for by a Markoff process, since a proportion of greater than one of D_{p1} would have been needed to account for the serial correlation. A second order process fitted using the third to sixth variate differences gave $A = -1{\cdot}03$, $B = 0{\cdot}38$, while one using the second to sixth variate differences gave $A = -1{\cdot}02$, $B = 0{\cdot}39$. The latter values correspond to $a = -0{\cdot}55$, $b = 0{\cdot}27$, which is not unlike a moving average process for which $\rho_1 = 0{\cdot}44$. The following table gives the values of V_i for this process.

	0	1st	2nd	3rd	4th	5th
1	851·388	40·775	91·896	147·772	116·311	232·453
2	837·796	−14·363	−13·655	57·308	−73·878	51·879
3	840·190	−8·900	−44·356	103·275	−109·250	80·180
4	841·080	3·773	−87·388	158·894	−155·507	—
5	840·828	22·499	−140·352	224·867	—	—
6	839·757	45·891	−201·680	—	—	—
7	838·118	72·781	—	—	—	—
8	836·096	—	—	—	—	—

There is apparently a low serial correlation in the revised series with a trend component of about 18 in the first variate difference.

10. DISCUSSION

Since it is theoretically possible to reproduce any symmetrical derived second-moment statistics using the quantities V_i, results obtainable using any method of general trend elimination should be capable of reproduction using some modification of the variate-difference method. For example, the use of symmetrical moving average difference formulae is equivalent to using a smoothed second difference and, therefore, any results obtained by this method should be obtainable using some modification to the variate difference method.

Of course, as has been shown (Spencer-Smith, 1947; Quenouille, 1949), the conventional methods of eliminating trend are liable to lead to biased conclusions concerning the nature of the series, and, therefore, any alternative method which was unbiased or biased to a lesser degree would seem to have a claim for consideration.

The analysis described above is certainly more efficient for random series, while, for non-random series, it seems likely to be both less biased and more efficient than the usual method. It also has the added advantage of, in effect, reserving judgement on the extent and complexity of the trend until the serial correlation can be examined simultaneously. It cannot be denied that it requires fairly lengthy calculations, but these would not seem to be greatly more than the conventional analyses now employed and would seem more amenable to modern calculating equipment, whereby differencing and summing of squares can be carried out.

At this stage, many more practical problems will have to be tackled before the utility or otherwise of this method is established. It is interesting, however, to note that the old and, in many ways, discredited variate-difference method can by some slight modifications overcome many of the criticisms to which it is subject.

APPENDIX

It is necessary to consider the effect of extrapolation on

$$V_i = V_0\left[1-\frac{2i}{i+1}r_1+\frac{2i(i-1)}{(i+1)(i+2)}r_2-\dots\right]$$

when this is known for $i = m+1, \dots, m+p$.

It is not hard to show that if

$$f(i) = \frac{1}{a+i}-\frac{1}{a} \quad \text{for} \quad i = m+1, \dots, m+p,$$

then the value of $f(0)$ obtained by extrapolation is

$$-\frac{(m+1)(m+2)\dots(m+p)}{a(a+m+1)(a+m+2)\dots(a+m+p)}.$$

This follows, since
$$\nabla^j f(m+1) = \frac{j!}{(a+m+1)(a+m+2)\dots(a+m+j+1)}$$

and
$$f(0) = f(m+1)+(m+1)\nabla f(m+1)+\frac{(m+1)(m+2)}{2!}\nabla^2 f(m+1)+\dots.$$

As a corollary of this result, the extrapolate to zero may be found for any function which may be split into partial fractions. Thus the extrapolate to zero for the function

$$f(i) = \frac{i(i-1)\dots(i-q+1)}{(i+1)(i+2)\dots(i+q)} \quad \text{for} \quad i = m+1, \dots, m+p$$

equals

$$\sum_{k=1}^{q}\frac{(-1)^{k-1}(q+k-1)!}{(q-k)!\,k!\,(k-1)!}\frac{(m+k)!}{m!}\frac{(m+p)!}{(m+p+k)!} = \sum_{k=1}^{q}(-1)^{k-1}\binom{q-1}{k-1}\frac{(q+k-1)!\,(m+k)!\,(m+p)!}{(m+p+k)!\,m!\,k!\,(q-1)!}.$$

In this latter form it may be seen to be the $(q-1)$th difference of the function $\dfrac{(q+k-1)!\,(m+k)!\,(m+p)!}{(m+p+k)!\,m!\,k!\,(q-1)!}$.

If $q > m+p$, this function is a polynomial of order $q-p$ and, consequently (as might be expected), the extrapolate is zero.

If $q \leqslant m+p$, it may be noted that

$$\frac{(m+k)!}{k!\,(m+p+k)!} = \frac{1}{(p+k)!}-\frac{\binom{m}{1}p}{(p+k+1)!}+\frac{\binom{m}{2}p(p+1)}{(p+k+2)!}-\dots,$$

and that the $(q-1)$th difference of $\dfrac{(q+k-1)!}{(p+k+a)!}$ is $(-1)^{q-1}\dfrac{(a+p-1)!}{(a+p-q)!}\dfrac{q!}{(a+p+q)!}$. Using these two facts, the extrapolate may be put into the alternative form

$$(-1)^{q-1}\frac{q(m+p)!\,(p-1)!}{m!(p-q)!\,(p+q)!}\left[1-\frac{\binom{m}{1}p^2}{[(p+1)^2-q^2]}+\frac{\binom{m}{2}p^2(p+1)^2}{[(p+1)^2-q^2][(p+2)-q^2]}-\dots\right].$$

The proof may be completed by noting that this expression is of order $1/p$, when m is zero and p is large, q being regarded as small, and that even when q is of the order of p ($= Xp$, say) the expression is of the order $(1+X)^{-p}$ at most. For other values of m, a similar result may be obtained by considering the modification to the expression for $m = 0$.

REFERENCES

Anderson, O. (1914). *Biometrika*, **10**, 269.

Bartlett, M. S. (1935). *J. R. Statist. Soc.* **98**, 536.

Davis, H. T. (1941). *The Analysis of Economic Time Series*. Bloomington, Indiana: Principia Press Inc.

Kendall, M. G. (1943). *J. R. Statist. Soc.* **106**, 91.

Kendall, M. G. (1945). *J. R. Statist. Soc.* **108**, 93.

Kendall, M. G. (1946). *Contributions to the Study of Oscillatory Time-Series*. Cambridge: University Press.

Kendall, M. G. (1949). *Biometrika*, **36**, 267.

Quenouille, M. H. (1949). *Travaux présentés à la 26ᵉ Session de l'Institut international de Statistique, Berne*, 1949. (*Bull. Inst. Int. Statist.* **32**, 290.)

Quenouille, M. H. (1952). *Associated Measurements*. London: Butterworths.

Spencer-Smith, J. L. (1947). *J. R. Statist. Soc.* (B), **9**, 104.

MOMENTS OF THE RANK CORRELATION COEFFICIENT τ IN THE GENERAL CASE*

By R. M. SUNDRUM

Division of Research Techniques, London School of Economics

1. Introduction

The mean and variance of τ are known in the general case and for a number of special cases. Some of these results are given below in terms of p, the probability of concordance. Two observations (x_i, y_i) and (x_j, y_j) are said to be concordant if $(x_i - x_j)(y_i - y_j) > 0$; then it can be seen easily that p is related to τ by the formula

$$\tau = 2p - 1 = p - q,$$

where $q = 1 - p$.

(a) *General case*

If we write p' for the sample estimate of p (and similarly use a prime to distinguish sample statistics from their corresponding parental values), we have the result of Hoeffding (1947) for the general case:

$$E(p') = p, \tag{1}$$

$$\sigma^2(p') = \frac{2}{n(n-1)}\{pq + 2(n-2)(k-p^2)\}, \tag{2}$$

where k is the probability that out of three members drawn from the population, one is concordant with the other two.

(b) *Special case: independence*

For this case, Kendall (1938) has obtained

$$E(p') = \tfrac{1}{2}, \tag{3}$$

$$\sigma^2(p') = \frac{2n+5}{18n(n-1)}. \tag{4}$$

(c) *Special case: normal correlation*

For this case, the mean and variance, due to Greiner and Esscher respectively, are given by Kendall (1948):

$$E(p') = \frac{1}{2} + \frac{1}{\pi}\sin^{-1}\rho, \tag{5}$$

$$\sigma^2(p') = \frac{1}{2n(n-1)}\left[1 - \left(\frac{2}{\pi}\sin^{-1}\rho\right)^2 + 2(n-2)\left\{\frac{1}{9} - \left(\frac{2}{\pi}\sin^{-1}\tfrac{1}{2}\rho\right)^2\right\}\right]. \tag{6}$$

Further, Hoeffding (1947) and Daniels & Kendall (1947) have proved that the distribution of t is asymptotically normal in the general case. Therefore, in the case of large samples, the above results are sufficient to fix the distribution completely. But it is notorious that the distribution of correlation coefficients in the non-null case tends to be extremely skew, particularly because of the limited range from -1 to $+1$; and that a very large sample is required before the skewness of the distribution is small enough to be neglected. The case of medium-size samples, which are likely to be most important in practice, therefore requires some consideration of the departure from normality of this sampling distribution. The third and fourth moments of this distribution are known only for the null case of independence. The cumulant-generating function of the distribution in this case was derived by Moran (1950) and Silverstone (1950), and from their result we obtain

---

* Part of a thesis submitted for the degree of Ph.D. in the University of London.

$$\mu_3(p') = 0, \tag{7}$$

$$\mu_4(p') = \frac{100n^4+328n^3-127n^2-997n-372}{2700n^3(n-1)^3}. \tag{8}$$

In this paper, formulae for the third and fourth moments are derived for the general case, and for a special case of normal correlation. But first we prove a conjecture of Daniels & Kendall (1947) about the upper bound of the variance.

2. UPPER BOUND OF THE VARIANCE OF t

Daniels & Kendall (1947) considered a population of ranked bivariate observations in which the sampling variance of t was maximized. They surmised that this occurred when the parent ranking was of the 'canonical' form, i.e. an arrangement where the X's are in the natural order, a certain number of the Y's at one end or the other of the sequence were in the correct natural order, and the others in the inverse order. A proof of this result is given below, in terms of p.

Let the parent population consist of N bivariate observations, such that the probability of concordance is p; this means that out of the $N(N-1)$ pairs (each pair counted twice) there are $P = pN(N-1)$ concordant pairs. Write P_i for the number of observations concordant with the member whose X-rank is i; then $P = \sum_i P_i$. The problem is then to find a subdivision of a given P into a set of P_i's such that the sampling variance of p' is a maximum. From (2) above, we see that in order to maximize the variance of p', for a given p, we have to maximize k.

Take two quantities P_i and P_j such that $P_i \geqslant P_j$. The contribution to k from these two is clearly $P_i(P_i-1)$ and $P_j(P_j-1)$. Now consider the effect of reducing P_j by one and increasing P_i by one. The contribution to k is then

$$P_i(P_i+1)+(P_j-1)(P_j-2), \quad \text{i.e.} \quad P_i(P_i-1)+P_j(P_j-1)+2(P_i-P_j)+2,$$

which is always greater than the original contribution; therefore the effect of such an alteration is always to increase the value of k. Therefore, k is maximized by increasing as many P_i as possible to their maximum value and reducing the other P_i as much as possible.

Now the maximum number of concordances with any single observation is clearly $(N-1)$, and for every observation with this number of concordances there is a contribution of 1 to every other P_i. Therefore the least possible value of P_i is equal to the number of observations with $(N-1)$ concordances. Further, in a ranking such as the following,

X-ranks	1	2	3	...	N
Y-ranks	Y_1	Y_2	Y_3	...	Y_N
Concordances	P_1	P_2	P_3	...	P_N

for any $P_i = (N-1)$, the rank Y_i must be the same as the X-rank i; all Y's to the left must be less than Y_i, and all Y's to the right must be greater than Y_i. Therefore all observations to the left of (i, Y_i) must be concordant with the observations to the right of these, and so these concordances are not reduced to their least possible value. In order to do this, the maximum concordances must occur at one or the other end of the sequence. Therefore the ranking which maximizes k is one in which, for a given P, as many P_i as possible at one end are assigned their maximum value $(N-1)$ and the remainder are assigned their minimum value equal to the number of observations with $(N-1)$ concordances. It can be shown easily that this leads to a canonical ranking. This can be done in all cases, with the possible exception of one observation.

3. THIRD AND FOURTH MOMENTS IN THE GENERAL CASE

To find
$$\mu_3(p') = E(p')^3 - 3E(p')^2\, E(p') + 2E(p')^3$$
we first consider $(p')^3$. Arguing in the manner of Hoeffding (1947), $(p')^3$ is the probability of three pairs drawn from the sample, with replacement, being concordant. We may express this probability in terms of the following probabilities involving members of the sample drawn without replacement. Then $(p')^3$ is the sum of the following eight probabilities:

(i) The probability of drawing the same pair three times, ${\binom{n}{2}}^{-2}$, multiplied by the probability that it is concordant, p':
$${\binom{n}{2}}^{-2} p'.$$

(ii) The probability that two pairs are identical, but the third has one member in common with them, $3{\binom{n}{2}}^{-2}\,.\,2(n-2)$, multiplied by the probability, termed k' by Hoeffding that among three members drawn from a sample, one is concordant with the other two:
$$6{\binom{n}{2}}^{-2}(n-2)\,k'.$$

(iii) The probability that two pairs are identical, while the third is distinct, $3{\binom{n}{2}}^{-2}\binom{n-2}{2}$, multiplied by the probability of drawing two concordant pairs, without replacement, i.e. $(p^2)'$, since the corresponding probability for the population is (p^2):
$$3{\binom{n}{2}}^{-2}\binom{n-2}{2}(p^2)'.$$

(iv) The probability that a pair has one member in common with the second pair, and the other member in common with the third, $3{\binom{n}{2}}^{-2}2(n-2)(n-3)$, multiplied by the probability, l', that in a sample of four, two concordant members can be chosen in such a way that each of them is concordant with one of the remaining two:
$$6{\binom{n}{2}}^{-2}(n-2)(n-3)\,l'.$$

(v) The probability that two pairs have one member in common, while the third is completely distinct, $3{\binom{n}{2}}^{-2}2(n-2)\binom{n-3}{2}$, multiplied by the probability, $(kp)'$, i.e. the probability of getting a set of three members with the property mentioned in (ii) above, multiplied by the probability of a distinct concordant pair:

$$6\binom{n}{2}^{-2}\binom{n-3}{2}(n-2)(kp)'.$$

(vi) The probability of drawing three pairs, all with one member in common, and with the other members of each pair distinct, $2\binom{n}{2}^{-2}(n-2)(n-3)$, multiplied by the probability t' that among four members drawn from the sample without replacement, one is concordant with the other three:

$$2\binom{n}{2}^{-2}(n-2)(n-3)t'.$$

(vii) The probability of drawing three pairs, any two of which have only one member in common, i.e. if there are three members, A, B, C, the three pairs are AB, AC, BC, $2\binom{n}{2}^{-2}(n-2)$, multiplied by the probability that among three members drawn from the sample without replacement, each is concordant with the other two, u':

$$2\binom{n}{2}^{-2}(n-2)u'.$$

(viii) The probability of drawing three distinct pairs, $\binom{n}{2}^{-2}\binom{n-2}{2}\binom{n-4}{2}$, multiplied by the probability that they are all concordant, $(p^3)'$:

$$\binom{n}{2}^{-2}\binom{n-2}{2}\binom{n-4}{2}(p^3)'.$$

In a similar manner, $(p')^4$, the probability of drawing four concordant pairs from a sample, replacing each pair, is the sum of the following twenty-three probabilities.

(i) The probability of drawing the same pair four times, $\binom{n}{2}^{-3}$, multiplied by the probability that it is concordant, p':

$$\binom{n}{2}^{-3}p'.$$

(ii) The probability of drawing three identical pairs, and the fourth with one in common, $4\binom{n}{2}^{-3}2(n-2)$, multiplied by the probability, k', that among three members drawn from the sample, one is concordant with the other two:

$$8\binom{n}{2}^{-3}(n-2)k'.$$

(iii) The probability of drawing three identical pairs, and the fourth distinct, $4\binom{n}{2}^{-3}\binom{n-2}{2}$, multiplied by the probability that two pairs drawn without replacement are concordant, $(p^2)'$:

$$2\binom{n}{2}^{-3}(n-2)(n-3)(p^2)'.$$

(iv) The probability of drawing two sets of identical pairs, with one in common, $3\binom{n}{2}^{-3} 2(n-2)$, multiplied by the probability, k':

$$6\binom{n}{2}^{-3}(n-2)\,k'.$$

(v) The probability of drawing two identical pairs, and the third and fourth with the same member in common with the first two, and other members distinct, $6\binom{n}{2}^{-3} 2(n-2)(n-3)$, multiplied by the probability, t', that among four members drawn without replacement from a sample, one is concordant with the other three:

$$12\binom{n}{2}^{-3}(n-2)(n-3)\,t'.$$

(vi) The probability of drawing two identical pairs, the third with one member in common with the first two, and the other member in common with the fourth pair, $6\binom{n}{2}^{-3} 4(n-2)(n-3)$, multiplied by the probability, l', that among four members drawn from a sample, two concordant members can be chosen in such a way that each of them is concordant with one of the remaining two:

$$24\binom{n}{2}^{-3}(n-2)(n-3)\,l'.$$

(vii) The probability of drawing two identical pairs, with one member in common with the third pair, and the other member in common with the fourth pair, $6\binom{n}{2}^{-3} 2(n-2)(n-3)$, multiplied by the probability, l' (as in (vi)):

$$12\binom{n}{2}^{-3}(n-2)(n-3)\,l'.$$

(viii) The probability of drawing two identical pairs, with one member in common with third pair, and a fourth pair distinct, $12\binom{n}{2}^{-3} 2(n-2)\binom{n-3}{2}$, multiplied by the probability, $(kp)'$, as in (i) and (ii):

$$12\binom{n}{2}^{-3}(n-2)(n-3)(n-4)\,(kp)'.$$

(ix) The probability of drawing two identical pairs, with the same member in common with the third and fourth pairs, whose other members are common, $6\binom{n}{2}^{-3} 2(n-2)$, multiplied by the probability, u', that among three members drawn from the sample without replacement, each is concordant with the other two:

$$12\binom{n}{2}^{-3}(n-2)\,u'.$$

(x) The probability of drawing each of two distinct pairs twice, $3\binom{n}{2}^{-3}\binom{n-2}{2}$, multiplied by the probability, $(p^2)'$, that two pairs drawn without replacement are concordant:

$$\frac{3}{2}\binom{n}{2}^{-3}(n-2)(n-3)(p^2)'.$$

(xi) The probability of drawing two identical pairs, and two other pairs with one in common, $6\binom{n}{2}^{-3}\binom{n-2}{2}2(n-4)$, multiplied by the probability, $(kp)'$, as in (viii):

$$6\binom{n}{2}^{-3}(n-2)(n-3)(n-4)(kp)'.$$

(xii) The probability of drawing two identical pairs, and two distinct pairs,

$$6\binom{n}{2}^{-3}\binom{n-2}{2}\binom{n-4}{2},$$

multiplied by the probability, $(p^3)'$, of getting three concordant pairs without replacement:

$$\frac{3}{2}\binom{n}{2}^{-3}(n-2)(n-3)(n-4)(n-5)(p^3)'.$$

(xiii) The probability of drawing three pairs with one member in common and others distinct, and the fourth pair having the uncommon members of the others,

$$2\binom{n}{2}^{-3}4(n-2)(n-3)\,.\,3,$$

multiplied by the probability, a', that among four members drawn from the sample without replacement, one is concordant with the other three, two of which are concordant with each other:

$$24\binom{n}{2}^{-3}(n-2)(n-3)\,a'.$$

(xiv) The probability of drawing two distinct pairs, the third pair having one member each from the first two pairs; and the fourth pair having the other members, $3\binom{n}{2}^{-3}4\binom{n-2}{2}$, multiplied by the probability, b', that among four members drawn from a sample without replacement, each is concordant with two other members:

$$6\binom{n}{2}^{-3}(n-2)(n-3)\,b'.$$

(xv) The probability of drawing three pairs with one in common, the fourth pair having one member in common with the uncommon members of the first three pairs, and another member distinct, $4\binom{n}{2}^{-3}2(n-2)(n-3)(n-4)$, multiplied by the probability, w', that among five members drawn without replacement, one is concordant with three others, one of which is concordant with the fifth:

$$24\binom{n}{2}^{-3}(n-2)(n-3)(n-4)\,w'.$$

(xvi) The probability of drawing a pair with one member in common with a second pair, and other member in common with a third pair, while the fourth pair has an uncommon member of the second or third, and a distinct member, $6\binom{n}{2}^{-3} 4(n-2)(n-3)(2n-4)$, multiplied by the probability, x', that among five members drawn without replacement from a sample, one is concordant with a second, which is concordant with a third, which is concordant with a fourth, which is concordant with the fifth:

$$24\binom{n}{2}^{-3}(n-2)(n-3)(n-4)x'.$$

(xvii) The probability of drawing one pair with one member in common with a second pair and other member in common with third pair, while a fourth pair is distinct,

$$6\binom{n}{2}^{-3} 4(n-2)(n-3)\binom{n-4}{2},$$

multiplied by the probability, $(lp)'$, as in (vi) and (i):

$$12\binom{n}{2}^{-3}(n-2)(n-3)(n-4)(n-5)(lp)'.$$

(xviii) The probability of drawing three pairs, any two of which have a member in common, while a fourth pair is distinct, $4\binom{n}{2}^{-3} 2(n-2)\binom{n-3}{2}$, multiplied by the probability, $(up)'$, as in (ix) and (i):

$$4\binom{n}{2}^{-3}(n-2)(n-3)(n-4)(up)'.$$

(xix) The probability of drawing three distinct pairs, one of which has a member in common with the fourth pair, $12\binom{n}{2}^{-3}\binom{n-2}{2}\binom{n-4}{2}(n-6)$, multiplied by the probability, $(kp^2)'$, as in (ii) and (iii):

$$3\binom{n}{2}^{-3}(n-2)(n-3)(n-4)(n-5)(n-6)(kp^2)'.$$

(xx) The probability of drawing four pairs with one member in common and other members distinct, $2\binom{n}{2}^{-3}(n-2)(n-3)(n-4)$, multiplied by the probability, y', that among five members drawn from a sample without replacement, one is concordant with other four:

$$2\binom{n}{2}^{-3}(n-2)(n-3)(n-4)y'.$$

(xxi) The probability of drawing three pairs with one member in common, and a fourth pair distinct, $4\binom{n}{2}^{-3} 2(n-2)(n-3)\binom{n-4}{2}$, multiplied by the probability, $(tp)'$, as in (v) and (i):

$$4\binom{n}{2}^{-3}(n-2)(n-3)(n-4)(n-5)(tp)'.$$

(xxii) The probability of drawing two pairs with one member in common, and two other pairs with a member in common, $3\binom{n}{2}^{-3}2(n-2)\binom{n-3}{2}2(n-5)$, multiplied by the probability, $(k^2)'$, as in (ii):

$$6\binom{n}{2}^{-3}(n-2)(n-3)(n-4)(n-5)(k^2)'.$$

(xxiii) The probability of drawing four distinct pairs, $\binom{n}{2}^{-3}\binom{n-2}{2}\binom{n-4}{2}\binom{n-6}{2}$, multiplied by the probability $(p^4)'$, of drawing four concordant pairs without replacement:

$$\frac{1}{128}\binom{n}{2}^{-3}(n-2)(n-3)(n-4)(n-5)(n-6)(n-7)(p^4)'.$$

Collecting these terms together, we get the following results:

$$\begin{aligned}\mu_3'(p') = \binom{n}{2}^{-2}\{&p+(n-2)(6k+2u)+(n-2)(n-3)(p^2+6l+2t)\\ &+(n-2)(n-3)(n-4)\,3kp+\tfrac{1}{4}(n-2)(n-3)(n-4)(n-5)\,p^3\},\end{aligned} \tag{9}$$

$$\begin{aligned}\mu_3(p') = \binom{n}{2}^{-2}\{&(6l+2t-18kp+10p^3)\,n^2+(6k+2u-6p^2-30l-10t+72kp-34p^3)\,n\\ &+(p+9p^2+30p^3-12k-4u+36l+12t-72kp)\},\end{aligned} \tag{10}$$

$$\begin{aligned}\mu_4'(p') = \binom{n}{2}^{-3}\{&p+(n-2)(14k+12u)\\ &+(n-2)(n-3)(\tfrac{7}{2}p^2+12t+36l+24a+6b)\\ &+(n-2)(n-3)(n-4)(18kp+24w+24x+4up+2y)\\ &+(n-2)(n-3)(n-4)(n-5)(\tfrac{3}{2}p^3+12lp+4tp+6k^2)\\ &+(n-2)(n-3)(n-4)(n-5)(n-6)\,3kp^2\\ &+\tfrac{1}{8}(n-2)(n-3)(n-4)(n-5)(n-6)(n-7)\,p^4\},\end{aligned} \tag{11}$$

$$\begin{aligned}\mu_4(p') = \binom{n}{2}^{-3}\{&6(k-p^2)^2\,n^4\\ &+(6kp+24w+24x+2y-6p^3-96lp-32tp-84k^2+270kp^2-108p^4)\,n^3\\ &+(\tfrac{3}{2}p^2+12t+36l+24a+6b-126kp-216w-216x\\ &\quad-24up-18y+75p^3+720lp+240tp+426k^2-1446kp^2+\tfrac{1011}{2}p^4)\,n^2\\ &+(14k+12u-\tfrac{31}{2}p^2-60t-180l-120a-30b\\ &\quad+444kp+624w+624x+96up+52y-213p^3\\ &\quad-1776lp-592tp-924k^2+2988kp^2-\tfrac{1887}{2}p^4)\,n\\ &+(p-28k-24u+21p^2+72t+216l+144a+36b-432kp-576w\\ &\quad-576x-96up-48y+180p^3+1440lp+480tp+720k^2\\ &\quad-2160kp^2+630p^4)\}.\end{aligned} \tag{12}$$

Considering only the dominant terms, we find

$$\sigma^2(p') \sim 4(k-p^2)\,n^{-1}, \tag{13}$$

$$\mu_3(p') \sim \text{const.}\, n^{-2}, \tag{14}$$

$$\mu_4(p') \sim 48(k-p^2)^2\,n^{-2}, \tag{15}$$

so that $\beta_1 \to 0$ and $\beta_2 \to 3$ as $n \to \infty$.

This agrees therefore with the general result for higher moments given by Hoeffding (1947) in his proof of the asymptotic normality of the distribution in the non-null case.

These formulae for the third and fourth moments involve ten parameters, namely, p, k, l, t, u, a, b, w, x and y. In the general case when nothing is known about the parent population, we have no information about these parameters, except that provided by the sample itself. Then we can use sample estimates of these quantities, suitably corrected for bias where necessary. These computations may, however, be much too tedious for common applications. It is therefore interesting to study the values of these parameters in special cases.

4. Special case of independence

We first evaluate the ten parameters in the null case of independence, partly because it provides a check on our algebra, and partly because it is useful to have these values in extending to the non-null case.

First we collect some preliminary results. Let X_i $(i = 1, 2, \ldots)$ be a series of random variables, independently and identically distributed according to some common distribution function $F(x)$. Define $u_{ij} = X_i - X_j$. Then we need the values of quantities like

$$E(\operatorname{sgn} u_{12} \operatorname{sgn} u_{13}),$$

where

$$\begin{aligned} \operatorname{sgn} u_{ij} &= 1 \quad \text{if} \quad u_{ij} > 0, \\ &= 0 \quad \text{if} \quad u_{ij} = 0, \\ &= -1 \quad \text{if} \quad u_{ij} < 0. \end{aligned}$$

We consider also $v_{ij} = Y_i - Y_j$, where Y_i $(i = 1, 2, \ldots)$ is a series of independently and identically distributed random variables.

Obviously, $E(\operatorname{sgn} u_{ij}) = 0$. Given X_1, the probability that both X_2 and X_3 are less than X_1 is $\{F(x_1)\}^2$; and the probability that both X_2 and X_3 are greater than X_1 is $\{1 - F(x_1)\}^2$. Therefore

$$\Pr.\{(X_1 - X_2)(X_1 - X_3) > 0\} = \int_0^1 \{F(x)\}^2\, dF(x) + \int_0^1 \{1 - F(x)\}^2\, dF(x) = \tfrac{2}{3},$$

$$\Pr.\{(X_1 - X_2)(X_1 - X_3) < 0\} = \tfrac{1}{3},$$

so that

$$E(\operatorname{sgn} u_{12} \operatorname{sgn} u_{13}) = \tfrac{1}{3}.$$

We can express the ten parameters in terms of the u's and the v's. Under the null hypothesis, every u is independent of any v. Hence, we can evaluate all these parameters. As an example, we evaluate the quantity k by this method:

$$\begin{aligned} k &= E\{\tfrac{1}{2}(\operatorname{sgn} u_{12} \operatorname{sgn} v_{12} + 1)\, \tfrac{1}{2}(\operatorname{sgn} u_{13} \operatorname{sgn} v_{13} + 1)\} \\ &= \tfrac{1}{4} E(\operatorname{sgn} u_{12} \operatorname{sgn} u_{13})\, E(\operatorname{sgn} v_{12} \operatorname{sgn} v_{13}) \\ &\quad + \tfrac{1}{4}\{E(\operatorname{sgn} u_{12})\, E(\operatorname{sgn} v_{12}) + E(\operatorname{sgn} u_{13})\, E(\operatorname{sgn} v_{13})\} \\ &\quad + \tfrac{1}{4} = \tfrac{5}{18}. \end{aligned}$$

In this way, we obtain for the null case:

$$\left. \begin{array}{lll} p = \frac{1}{2}, & u = \frac{1}{6}, & x = \frac{76}{900}, \\ k = \frac{5}{18}, & a = \frac{7}{72}, & y = \frac{8}{75}. \\ l = \frac{11}{72}, & b = \frac{7}{72}, & \\ t = \frac{1}{6}, & w = \frac{163}{1800}, & \end{array} \right\} \tag{16}$$

However, this method of evaluating the parameters does not lend itself simply to generalization to the non-null case. We therefore consider an alternative procedure used by Hoeffding (1947) to evaluate k, and based on the consideration that in the null case, all permutations of one set of ranks relative to the other are equally probable. In the following, we write $P(1423)$ to indicate the probability that the ranks of the Y's occur in that order when the X's are arranged in order of magnitude.

(i) $p = \frac{1}{2}$,

(ii) $k = P(123) + \frac{1}{3}P\{(132)+(213)\} = \frac{5}{18}$,

(iii)
$$\begin{aligned} l = {} & P(1234) + \tfrac{1}{2}P\{(1243)+(1324)+(2134)\} \\ & + \tfrac{1}{3}P(2143) + \tfrac{1}{6}P\{(1342)+(1423)+(2314)+(2413)+(3124)\} \\ = {} & \tfrac{11}{72}, \end{aligned}$$

(iv)
$$\begin{aligned} t = {} & P(1234) + \tfrac{1}{2}P\{(1243)+(1324)+(2134)\} \\ & + \tfrac{1}{4}P\{(1342)+(1423)+(1432)+(2314)+(3124)+(3214)\} \\ = {} & \tfrac{1}{6}, \end{aligned}$$

(v) $u = P(123) = \frac{1}{6}$,

(vi)
$$\begin{aligned} a = {} & P(1234) + \tfrac{1}{3}P\{(1243)+(1324)+(2134)\} \\ & + \tfrac{1}{12}P\{(1342)+(1423)+(2314)+(3124)\} \\ = {} & \tfrac{7}{72}, \end{aligned}$$

(vii)
$$\begin{aligned} b = {} & P(1234) + \tfrac{1}{3}P\{(1243)+(1324)+(2134)+(2143)\} \\ = {} & \tfrac{7}{72}, \end{aligned}$$

(viii)
$$\begin{aligned} w = {} & P(12345) + \tfrac{3}{5}P\{(12354)+(12435)+(13245)+(21345)\} \\ & + \tfrac{1}{3}P\{(12453)+(12534)+(13254)+(13425)+(14235) \\ & \quad +(21354)+(21435)+(23145)+(31245)\} \\ & + \tfrac{1}{5}P\{(12543)+(14325)+(32145)\} \\ & + \tfrac{3}{20}P\{(13452)+(15234)+(23415)+(41235)\} \\ & + \tfrac{1}{6}P\{(13524)+(14253)+(21453)+(21534)+(23154)+(24135) \\ & \quad +(31254)+(31425)\} \\ & + \tfrac{1}{12}P\{(13542)+(14352)+(15243)+(15324)+(24315) \\ & \quad +(32415)+(41325)+(42135)\} \\ & + \tfrac{1}{10}P\{(21543)+(32154)\} \\ & + \tfrac{1}{15}P\{(14523)+(23514)+(24153)+(25134)+(31452) \\ & \quad +(31524)+(34125)+(41253)\} \\ & + \tfrac{1}{30}P\{(14532)+(15342)+(15423)+(25143)+(25314)+(31542)+(32514) \\ & \quad +(34215)+(41352)+(42153)+(42315)+(43125)\} \\ & + \tfrac{1}{60}P\{(24513)+(25413)+(34152)+(35124)+(35214)+(41523)+(41532) \\ & \quad +(43152)\} \\ = {} & \tfrac{163}{1800}, \end{aligned}$$

$$\begin{aligned}
\text{(ix)}\quad x = {} & P(12345)+\tfrac{3}{5}P\{(12354)+(12435)+(13245)+(21345)\} \\
& +\tfrac{3}{10}P\{(12453)+(12534)+(13425)+(14235)+(23145)+(31245)\} \\
& +\tfrac{1}{10}P\{(12543)+(13452)+(14325)+(15234)+(21543)+(23415)+(32145) \\
& \quad +(32154)+(41235)\} \\
& +\tfrac{2}{5}P\{(13254)+(21354)+(21435)\} \\
& +\tfrac{1}{6}P\{(13524)+(14253)+(24135)+(31425)\} \\
& +\tfrac{1}{15}P\{(14523)+(23514)+(25134)+(31452)+(34125)+(41253)\} \\
& +\tfrac{7}{30}P\{(21453)+(21534)+(23154)+(31254)\} \\
& +\tfrac{7}{60}P\{(24153)+(31524)\} \\
& +\tfrac{1}{30}P\{(13542)+(14352)+(15243)+(15324)+(24315)+(24513)+(25143) \\
& \quad +(31542)+(32415)+(32514)+(34152)+(35124)+(41325)+(41523) \\
& \quad +(42135)+(42153)\} \\
& +\tfrac{1}{60}P\{(25314)+(35142)+(41352)+(42513)\} \\
= {} & \tfrac{76}{900},
\end{aligned}$$

$$\begin{aligned}
\text{(x)}\quad y = {} & P(12345)+\tfrac{3}{5}P\{(12354)+(12435)+(13245)+(21345)\} \\
& +\tfrac{2}{5}P\{(12453)+(12534)+(12543)+(13425)+(14235)+(14325)+(23145) \\
& \quad +(31245)+(32145)\} \\
& +\tfrac{1}{5}P\{(13254)+(13452)+(13524)+(13542)+(14253)+(14352) \\
& \quad +(14523)+(14532)+(15234)+(15243)+(15324) \\
& \quad +(15342)+(15423)+(15432)+(21354)+(21435)+(23415) \\
& \quad +(24135)+(24315)+(31425)+(32415)+(34125)+(34215) \\
& \quad +(41235)+(41325)+(42135)+(42315)+(43125)+(43215)\} \\
= {} & \tfrac{8}{75}.
\end{aligned}$$

When we substitute these values for the ten parameters in (10) and (12), we get the same result as (7) and (8).

5. SPECIAL CASE OF NORMAL CORRELATION

We now consider the values of the ten parameters in the case when the parent population is bivariate normal. For this case we already have, from (5)

$$p = \frac{1}{2}+\frac{1}{\pi}\sin^{-1}\rho.$$

Further, by comparing (2) and (6) we have

$$k = \frac{1}{4}\left\{\frac{10}{9}+\frac{4}{\pi}\sin^{-1}\rho+\frac{4}{\pi^2}(\sin^{-1}\rho)^2-\frac{4}{\pi^2}(\sin^{-1}\tfrac{1}{2}\rho)^2\right\}. \tag{17}$$

One other parameter can be evaluated exactly by the same methods used to deduce the above, i.e.

$$u = \frac{3}{8}\left\{\frac{4}{9}+\frac{4}{\pi}\sin^{-1}\rho-\frac{4}{\pi}\sin^{-1}\tfrac{1}{2}\rho+\frac{4}{\pi^2}(\sin^{-1}\rho)^2-\frac{4}{\pi^2}(\sin^{-1}\tfrac{1}{2}\rho)^2\right\}. \tag{18}$$

The other parameters cannot be determined exactly; their evaluation involves the same type of difficulty as that pointed out by Kendall (1949) in evaluating the variance of Spearman's rank correlation coefficient r_s exactly under normal correlation. However, estimates of these parameters can be obtained fairly easily, because from the results of the previous section we see that only samples of five need be considered. I have shown in another paper (Sundrum, 1953) how a systematic method of sampling may be used in such cases, and, in particular, used this method to obtain the frequency distribution of rankings of three, four and five from a bivariate normal population with correlation $\rho = 1/\sqrt{2}$. The theoretical values of three parameters (for the case $\rho = 1/\sqrt{2}$) and the sample estimates of the other seven parameters, obtained from the above sample results, are as follows:

$$\left.\begin{array}{ll} p = 0{\cdot}7500, & a = 0{\cdot}3470, \\ k = 0{\cdot}5770, & b = 0{\cdot}3485, \\ u = 0{\cdot}4431, & w = 0{\cdot}3451, \\ t = 0{\cdot}4543, & x = 0{\cdot}3363, \\ l = 0{\cdot}4414, & y = 0{\cdot}3642. \end{array}\right\} \qquad (19)$$

Substituting these values in (10) and (12), we get

$$\mu_3 = \binom{n}{2}^{-2}\{-0{\cdot}013625n^2+0{\cdot}0007n-0{\cdot}0427\}, \qquad (20)$$

$$\mu_4 = \binom{n}{2}^{-3}\{0{\cdot}0012615n^4-0{\cdot}04316n^3+0{\cdot}59027n^2-1{\cdot}9285n+1{\cdot}70314\}. \qquad (21)$$

These formulae are not entirely satisfactory, and are given here only as an illustration of the method. Sampling experiments to improve these results and to obtain corresponding formulae for a range of values of ρ have been put in hand.

However, the argument of this paper shows that the third and fourth moments of t depend only on ten parameters, and that these parameters may be evaluated by finding the probability distribution of rankings of five from correlated populations.

In conclusion, the author would like to express his indebtedness to Prof. M. G. Kendall for suggesting the problem and for many valuable suggestions during the preparation of this paper.

[*Note added in proof.*] With regard to the argument of section 2 on the upper bound of the variance of t, it has been objected (I am indebted to Dr Hartley for pointing this out to me) that the transfer of units between the quantities designated P_i and P_j is not always possible; i.e. that it has not been proved that it is always possible to find a ranking in which the P_i's have the altered values. This objection may be met as follows:

The problem is essentially one of maximizing a certain function of the P_i's, subject to $P=\sum_i P_i$, and certain other constraints on the P_i's by virtue of their definition, and the object of showing that there is a ranking corresponding to the altered values of the P_i's is to show that the set of P_i's satisfies *all* the constraints at *every* stage of the maximization procedure. In section 2 only some of these constraints are taken into account during the maximization procedure. However, it is shown that the maximizing set of P_i's *does* satisfy all the constraints because it corresponds to an actual ranking, i.e. the canonical ranking.

Now, a quantity which is maximized subject to only a part of the set of all constraints must be greater or equal to the maximum subject to all constraints. If, however, the solution so obtained satisfies all constraints, then it must be equal to the maximum subject to all constraints.

REFERENCES

Daniels, H. E. & Kendall, M. G. (1947). *Biometrika*, **34**, 197.
Hoeffding, W. (1947). *Biometrika*, **34**, 183.
Kendall, M. G. (1938). *Biometrika*, **30**, 81.
Kendall, M. G. (1948). *Rank Correlation Methods*. London: Charles Griffin.
Kendall, M. G. (1949). *Biometrika*, **36**, 177.
Moran, P. A. P. (1950). *J.R. Statist. Soc.* series B, **12**, 153.
Silverstone, H. (1950). *Biometrika*, **37**, 231.
Sundrum, R. M. (1953). *Biometrika*, **40**, 452.

99·9 AND 0·1 % POINTS OF THE χ^2 DISTRIBUTION

BY T. LEWIS, *Statistical Advisory Unit, Ministry of Supply*

1. In answer to a special inquiry the lower 0·1 % points of the χ^2 distribution have been computed correct to 8 decimal places for degrees of freedom $n = 1(1)\,30(10)\,100$, 120, and also the upper 0·1 % points for $n = 40(10)\,100$, 120. The resulting values which it is thought may be of value to other statisticians are given in Tables 1 and 2 below.

Table 1. *Lower 0·1 % points of χ^2*

Degrees of freedom	Value of χ^2	Degrees of freedom	Value of χ^2	Degrees of freedom	Value of χ^2
1	0·0000 0157	16	3·9416 2784	40	17·9164 2654
2	0·0020 0100	17	4·4160 9272	50	24·6739 0527
3	0·0242 9759	18	4·9048 4881	60	31·7383 4159
4	0·0908 0404	19	5·4068 1602	70	39·0363 7739
5	0·2102 1260	20	5·9210 4075	80	46·5198 7592
6	0·3810 6676	21	6·4466 7656	90	54·1552 4356
7	0·5984 9375	22	6·9829 6844	100	61·9179 3921
8	0·8571 0483	23	7·5292 3977	120	77·7551 4043
9	1·1519 4955	24	8·0848 8158		
10	1·4787 4346	25	8·6493 4363		
11	1·8338 5266	26	9·2221 2682		
12	2·2142 0932	27	9·8027 7692		
13	2·6172 1815	28	10·3908 7912		
14	3·0406 7252	29	10.9860 5349		
15	3·4826 8447	30	11·5879 5105		

Table 2. *Upper 0·1 % points of χ^2*

Degrees of freedom	Value of χ^2	Degrees of freedom	Value of χ^2
40	73·4019 5752	80	124·8392 2402
50	86·6608 1519	90	137·2083 5413
60	99·6072 3307	100	149·4492 5278
70	112·3169 3185	120	173·6174 3646

METHOD OF CALCULATION

2. We write n for degrees of freedom, $m \equiv \frac{1}{2}n$ and $v \equiv \frac{1}{2}\chi^2$. If, then, $2v$ is the $100P$ % point of the χ^2 distribution for n degrees of freedom, and $r!$ is written for $\Gamma(r+1)$ whether r is integral or not,

$$\frac{1}{\Gamma(m)}\int_v^\infty e^{-t}t^{m-1}\,dt = P,$$

or, in other words,

$$f(v) = 0,$$

where $$f(v) = \left\{1+v+\frac{v^2}{2!}+\ldots+\frac{v^{m-1}}{(m-1)!}\right\} - Pe^v \quad [n \text{ even}], \tag{1}$$

$$f(v) = \left\{\sqrt{\frac{2}{\pi}}\,e^v \int_{\sqrt{(2v)}}^{\infty} e^{-\frac{1}{2}t^2}dt + \frac{v^{\frac{1}{2}}}{(\frac{1}{2})!}+\frac{v^{\frac{3}{2}}}{(\frac{3}{2})!}+\ldots+\frac{v^{m-1}}{(m-1)!}\right\} - Pe^v \quad [n \text{ odd}]. \tag{2}$$

The exponential or quasi-exponential sum in curly brackets, i.e. $f(v)+Pe^v$, we write as S_{m-1} and the term $v^{m-1}/(m-1)!$ as T_{m-1}.

3. Each required value of $v(m, P)$ was obtained by computing $f(v)$ directly for an approximate value of v, $v = v_0$ say, and improving v_0 iteratively to $v_0+\delta v$ by means of the formula

$$\delta v = \lambda + \alpha\lambda^2 + \beta\lambda^3, \tag{3}$$

where
$$\left.\begin{aligned} \lambda &= \left.\frac{f(v)}{T_{m-1}}\right|_{v=v_0}, \\ \alpha &= \frac{1}{2}\left(\frac{T_{m-1}-T_{m-2}}{T_{m-1}}\right)_{v=v_0}, \\ \beta &= 2\alpha^2 - \frac{1}{6}\left(\frac{T_{m-1}-2T_{m-2}+T_{m-3}}{T_{m-1}}\right)_{v=v_0}. \end{aligned}\right\} \tag{4}$$

This formula for δv follows from the Taylor expansion of v as a function of the small quantity $w \equiv e^{-v}f(v)$, thus:

$v = v_0$ when $w = e^{-v_0}f(v_0)$, $v = v_0+\delta v$ when $w = 0$, hence

$$\delta v = -e^{-v_0}f(v_0)\left(\frac{dv}{dw}\right)_0 + \tfrac{1}{2}e^{-2v_0}f^2(v_0)\left(\frac{d^2v}{dw^2}\right)_0 - \tfrac{1}{6}e^{-3v_0}f^3(v_0)\left(\frac{d^3v}{dw^3}\right)_0 + \ldots.$$

But $w = e^{-v}S_{m-1} - P$, hence

$$\frac{dw}{dv} = e^{-v}(S_{m-2}-S_{m-1}) = -e^{-v}T_{m-1},$$

$$\frac{d^2w}{dv^2} = e^{-v}(T_{m-1}-T_{m-2}),$$

$$\frac{d^3w}{dv^3} = -e^{-v}(T_{m-1}-2T_{m-2}+T_{m-3}),$$

whence the values of the inverse derivatives d^rv/dw^r and the relations (3) and (4) readily follow.

OUTLINE OF PROCEDURE

4. For the first approximation v_0, use was made of the well-known formula

$$\sqrt[3]{(v/m)} \sim 1 - \frac{1}{9m} + x\sqrt{\frac{1}{9m}},$$

due to Wilson & Hilferty (1931), in which the $100P$ % value of v is related to the corresponding unit normal deviate x. The terms $T_r = v_0^r/r!$ were then computed in succession, starting from the greatest term T_q; the extra term

$$T_{-\frac{1}{2}} = \sqrt{\frac{2}{\pi}}\,e^{v_0}\int_{\sqrt{(2v_0)}}^{\infty} e^{-\frac{1}{2}t^2}\,dt$$

required for odd n was obtained separately from Sheppard's (1939) 12-decimal table of $e^{\frac{1}{2}x^2}\int_x^\infty e^{-\frac{1}{2}t^2}\,dt$ and its derivatives.

The terms T_r were then added together to give the sum S_{m-1} and also the last few partial sums $\ldots, S_{m-3}, S_{m-2}$. Using the National Bureau of Standards (1951) 15- and 18-decimal tables of $e^{\pm x}$, e^{v_0} was obtained to the requisite number of places; the difference

$$f(v_0) = S_{m-1} - Pe^{v_0}$$

was then taken, and the correction δv computed from the relations (3) and (4).

For the next iteration, with $v = v_0 + \delta v$, no computation of terms T_r was necessary to obtain the new sum $S_{m-1}(v_0 + \delta v)$, this being calculated with ease from the first few terms of the Taylor series

$$S_{m-1}(v_0+\delta v) = S_{m-1} + \delta v\, S_{m-2} + \tfrac{1}{2}\delta v^2 S_{m-3} + \tfrac{1}{6}\delta v^3 S_{m-4} + \ldots, \tag{5}$$

the S_r on the right-hand side being the values for $v = v_0$ already recorded. In the same way $T_{m-1}(v_0 + \delta v)$ was obtained from the Taylor series

$$T_{m-1} + \delta v\, T_{m-2} + \tfrac{1}{2}\delta v^2 T_{m-3} + \ldots.$$

After this, one further iteration at most was required, the sum $S_{m-1}(v_0 + \delta v + \delta v')$ being evaluated as above by using the accumulated correction $\delta v + \delta v'$ in the Taylor series (5).

ACCURACY OF METHOD

5. In the above process, v is determined as the value for which the 'discrepancy' $f(v) = S_{m-1} - Pe^v$ comes out to zero. If in fact $S_{m-1} - Pe^v$ differs from 0 by a small amount ϵ, the error $\delta_1 v$ in v (so written because opposite in sign to the δv of paras. 3 and 4), is given by

$$\begin{aligned}\epsilon &= \delta(S_{m-1} - Pe^v) = (S_{m-2} - Pe^v)\,\delta_1 v \\ &\sim (S_{m-2} - S_{m-1})\,\delta_1 v = -T_{m-1}\,\delta_1 v,\end{aligned}$$

i.e.

$$\delta_1 v \sim -\epsilon/T_{m-1}.$$

To estimate ϵ, let us suppose that the pivotal term T_q is calculated to N significant figures, 1 unit in the last place being denoted by u, so that $10^{-(N-1)}T_q > u > 10^{-N}T_q$. Clearly T_q and the few terms T_r of like order (i.e. with the same number of figures before the decimal point) have errors of order u, the terms of next lower order (with one less figure before the decimal point) have errors of order $10^{-1}u$, and so on; the error in Pe^v is negligible. Hence ϵ is at most of order $10^{-(N-1)}T_q$.

When $P = 0{\cdot}001$, $T_q \equiv T_{m-1}$, hence $|\delta_1 v|$ is of order $10^{-(N-1)}$, and v is certainly correct to $N-2$ decimal places.

When $P = 0{\cdot}999$,

$$T_q/T_{m-1} \sim e^{\frac{1}{2}x^2} = e^{\frac{1}{2}(3{\cdot}090)^2} \sim 10^2,$$

hence $|\delta_1 v|$ is of order

$$10^{-(N-1)}T_q/T_{m-1}, \quad \text{i.e.} \quad 10^{-(N-3)}.$$

Thus v is certainly correct to $N-4$ decimal places.

Accordingly the terms T were calculated to 12 significant figures for $P = 0{\cdot}999$ and to 10 significant figures for $P = 0{\cdot}001$.

CHECKING OF FINAL RESULTS BY DIFFERENCING

6. It can readily be shown that $\chi^2_n(P)$, the $100P$ % point of χ^2 for n degrees of freedom, is of the form

$$n + an^{\frac{1}{2}} + b + cn^{-\frac{1}{2}} + dn^{-1} + en^{-\frac{3}{2}} + \ldots, \tag{6}$$

where $a, b, c, \ldots$ are functions of P.

For this reason, direct differencing of the results $\chi^2_n(0{\cdot}999)$ against n for $n = 1(1)\,30$ and again for $n = 30(10)\,100$, and of $\chi^2_n(0{\cdot}001)$ for $n = 40(10)\,100$, was not practicable. An attempt to base the differencing check on the residuals

$$\{\chi^2_n(P) - \text{corresponding Wilson-Hilferty approximation}\}$$

also failed. In the end it was found possible to check all the results for $n \geqslant 14$ by a differencing method based on the accurate evaluation of series (6) as far as the term in $n^{-\frac{3}{2}}$, and to check the results for $n \leqslant 14$ by means of another series. Details of these series are given in paras. 7 and 8 respectively, and the steps in the checking procedure are outlined in para. 9.

7. *Series for* χ^2 *in terms of the corresponding normal deviate.* If x is the $100P$ % normal deviate, so that

$$\frac{1}{\sqrt{(2\pi)}}\int_x^\infty e^{-\frac{1}{2}t^2}\,dt = P = \frac{1}{\Gamma(\frac{1}{2}n)}\int_{\frac{1}{2}\chi^2}^\infty e^{-t}\,t^{\frac{1}{2}n-1}\,dt,$$

then the coefficients $a, b, c, \ldots$ in (6) can be expressed as polynomials in x. Developing the series for convenience in terms of $m \equiv \frac{1}{2}n$ instead of n, we may write

$$\chi^2 = 2m + a_0 x\, m^{\frac{1}{2}} + (b_0 x^2 + b_1) + (c_0 x^3 + c_1 x)\, m^{-\frac{1}{2}} + (d_0 x^4 + d_1 x^2 + d_2)\, m^{-1} + \ldots, \tag{7}$$

where $a_0, b_0, b_1, c_0, \ldots$ are constants. The values of these constants can be determined by equating moments about zero for the left- and right-hand sides of (7), bearing in mind that the rth moment of χ^2 is $2^r m(m+1)\ldots(m+r-1)$ and that of x, 0 when r is odd and $(r-1)(r-3)\ldots 3.1$ when r is even. For example, if we take (7) for convenience in the form

$$\chi^2 - 2m = a_0 x m^{\frac{1}{2}} + (b_0 x^2 + b_1) + O(m^{-\frac{1}{2}})$$

and equate 1st, 2nd and 3rd moments each side, we get

$$\begin{aligned} 2m - 2m &= b_0 + b_1 + O(m^{-1}),\\ 4m(m+1) - 4m\,.\,2m + 4m^2 &= a_0^2 m + O(1),\\ 8m(m+1)(m+2) - 6m\,.\,4m(m+1) + 12m^2\,.\,2m - 8m^3 &= 3a_0^2(3b_0 + b_1)\, m + O(1), \end{aligned}$$

whence $b_0 + b_1 = 0$, $a_0^2 = 4$, $3b_0 + b_1 = \frac{4}{3}$. Clearly the sign of a_0 is $+$, since $\chi^2 \gg 2m$ when $x \gg 0$; thus series (7) begins

$$\chi^2 = 2m + 2x m^{\frac{1}{2}} + \tfrac{2}{3}(x^2 - 1) + \ldots.$$

In this manner the expansion may be taken to any required number of terms; to order $m^{-\frac{3}{2}}$ it is

$$\begin{aligned} \chi^2 = 2m + 2x m^{\frac{1}{2}} + \tfrac{2}{3}(x^2 - 1) + \tfrac{1}{18}(x^3 - 7x)\, m^{-\frac{1}{2}}&\\ - \tfrac{1}{405}(3x^4 + 7x^2 - 16)\, m^{-1} + \tfrac{1}{19440}(9x^5 + 256x^3 - 433x)\, m^{-\frac{3}{2}} + \ldots.& \end{aligned} \tag{8}$$

In particular, when $P = 0{\cdot}001$, $x = +3{\cdot}09023,\,23062$,

and when $P = 0{\cdot}999$, $x = -3{\cdot}09023,\,23062$;

hence, from (8),

$$\begin{aligned} \chi^2_{2m}(0{\cdot}001) = 2m &+ 6{\cdot}18046,\,46124 m^{\frac{1}{2}} + 5{\cdot}69969,\,04708\\ &+ 0{\cdot}43770,\,32003 m^{-\frac{1}{2}} - 0{\cdot}80105,\,59174 m^{-1}\\ &+ 0{\cdot}45024,\,93635 m^{-\frac{3}{2}} + \ldots \end{aligned} \tag{9}$$

$$= 2m + u_{\frac{1}{2}} m^{\frac{1}{2}} + u_0 + u_{-\frac{1}{2}} m^{-\frac{1}{2}} + u_{-1} m^{-1} + u_{-\frac{3}{2}} m^{-\frac{3}{2}} + \ldots, \text{ say;} \tag{10}$$

and

$$\chi^2_{2m}(0{\cdot}999) = 2m - u_{\frac{1}{2}} m^{\frac{1}{2}} + u_0 - u_{-\frac{1}{2}} m^{-\frac{1}{2}} + u_{-1} m^{-1} - u_{-\frac{3}{2}} m^{-\frac{3}{2}} + \ldots. \tag{11}$$

8. *Check series for $\chi^2_n(P)$ when n is small and P nearly 1.* It is clear from (1) and (2) that when P is near to unity, so that v/m is small,

$$(1-P)\,e^v = \frac{v^m}{m!}+\frac{v^{m+1}}{(m+1)!}+\dots$$
$$= \frac{v^m}{m!}\left(1+\frac{v}{m+1}+\frac{v^2}{(m+1)(m+2)}+\dots\right);$$

hence
$$v^m = m!\,(1-P)\,e^v\left(1+\frac{v}{m+1}+\dots\right)^{-1}$$

or
$$v = [m!\,(1-P)]^{1/m}\,e^{v/m}\left(1+\frac{v}{m+1}+\dots\right)^{-1/m};$$

on expanding the exponential and binomial, multiplying together, and putting $2v = \chi^2$, this gives

$$\begin{aligned}\chi^2 = 2[m!(1-P)]^{1/m}\Bigg\{&1+\frac{\chi^2}{n+2}+\frac{(\chi^2)^2}{2(n+2)(n+4)}\\ &+\frac{(\chi^2)^3}{6(n+2)(n+4)(n+6)}\,\frac{n-2}{n+2}\\ &+\frac{(\chi^2)^4}{24(n+2)(n+4)(n+6)(n+8)}\,\frac{n^3-20n^2-124n-64}{(n+2)^2(n+4)}\\ &+\frac{(\chi^2)^5}{120(n+2)(n+4)(n+6)(n+8)(n+10)}\,\frac{(n-2)(n^3-136n^2-916n-704)}{(n+2)^3(n+4)}+\dots\Bigg\}.\end{aligned} \tag{12}$$

9. *Checking procedure.* The comparison between series (8) and the computed values of χ^2 furnished an adequate basis for a differencing check over the higher range of values of n, say $n \geqslant 30$, but not for n between say 15 and 30. On the other hand, the algebraic computation of further polynomial coefficients in (8) would have been laborious. The following method was therefore adopted:

Using equations (9), (10) and (11) and the set of paired results $\chi^2_n(0{\cdot}001)$, $\chi^2_n(0{\cdot}999)$ available for $n = 40(10)\,100$, the quantities

$$g_n \equiv \tfrac{1}{2}[\chi^2_n(0{\cdot}001)+\chi^2_n(0{\cdot}999)]-2m-u_0-u_{-1}m^{-1},$$
$$h_n \equiv \tfrac{1}{2}[\chi^2_n(0{\cdot}001)-\chi^2_n(0{\cdot}999)]-u_{\frac{1}{2}}m^{\frac{1}{2}}-u_{-\frac{1}{2}}m^{-\frac{1}{2}}-u_{-\frac{3}{2}}m^{-\frac{3}{2}},$$

were calculated, checked by differencing, and recomputed where necessary. Estimates $\hat{u}_{-2}, \hat{u}_{-3}, \hat{u}_{-4}$ of the coefficients of m^{-2}, m^{-3}, m^{-4} in (9) were then obtained by fitting the seven quantities g_n to the relationship

$$g_n = \hat{u}_{-2}m^{-2}+\hat{u}_{-3}m^{-3}+\hat{u}_{-4}m^{-4};$$

similarly, $\hat{u}_{-\frac{5}{2}}, \hat{u}_{-\frac{7}{2}}, \hat{u}_{-\frac{9}{2}}$ were obtained from the h_n.

The values of $\chi^2_n(0{\cdot}999)$ for $n = 14(1)\,30$ were now checked by differencing the residuals

$$\chi^2-(2m-u_{\frac{1}{2}}m^{\frac{1}{2}}+u_0-\dots-u_{-\frac{3}{2}}m^{-\frac{3}{2}}+\hat{u}_{-2}m^{-2}-\hat{u}_{-\frac{5}{2}}m^{-\frac{5}{2}}+\dots-\hat{u}_{-\frac{9}{2}}m^{-\frac{9}{2}}).$$

For $n = 1(1)\,14$ the values of $\chi^2_n(0{\cdot}999)$ were checked from equation (12) by evaluating for each given value of χ^2 the discrepancy between the two sides of the equation and differencing the quantities so obtained.

STANDARD APPROXIMATIONS TO χ^2 WHEN n IS LARGE

10. It is of interest to compare the two most familiar approximations for χ^2 with the exact series obtained above.

I. $\sqrt{(2\chi^2)}$ distributed approximately normally with mean $\sqrt{(2n-1)}$ and variance 1 (Fisher).

This gives $$\chi^2 = n + x\sqrt{2}\,n^{\frac{1}{2}} + \tfrac{1}{2}(x^2-1) + O(n^{-\frac{1}{2}}).$$

II. $\sqrt[3]{(\chi^2/n)}$ distributed approximately normally with mean $1-2/(9n)$ and variance $2/(9n)$ (Wilson-Hilferty).

This gives $$\chi^2 = n + x\sqrt{2}\,n^{\frac{1}{2}} + \tfrac{2}{3}(x^2-1) + \frac{2\sqrt{2}}{27}(x^3-6x)\,n^{-\frac{1}{2}} + O(n^{-1}).$$

III. Exact series (equation (8)):

$$\chi^2 = n + x\sqrt{2}\,n^{\frac{1}{2}} + \tfrac{2}{3}(x^2-1) + \frac{\sqrt{2}}{18}(x^3-7x)\,n^{-\frac{1}{2}} + O(n^{-1}).$$

I wish to thank Mr E. C. Fieller for his advice on method, and my colleagues in the Statistical Advisory Unit, Ministry of Supply, for their collaboration in the computing work. I also wish to make acknowledgement to the Chief Scientist, Ministry of Supply, for permission to publish this paper.

REFERENCES

SHEPPARD, W. F. (1939). *The Probability Integral.* Brit. Ass. Mathematical Tables, **7**. Cambridge University Press.

WILSON, E. B. & HILFERTY, M. M. (1931). *Proc. Nat. Acad. Sci., Wash.*, **17**, 684.

Tables of the Exponential Function e^x. (1951). National Bureau of Standards, Applied Mathematics Series, no. 14. New York.

TABLES OF SYMMETRIC FUNCTIONS. PART IV

BY F. N. DAVID AND M. G. KENDALL

1. In recent publications (David & Kendall, 1949, 1951) we have given, for functions up to and including weight 12, tables of the monomial symmetric functions (M) in terms of the one-part functions (S) and vice versa, tables of the unitary functions (U) in terms of M and vice versa, and tables of the homogeneous product sums (H) in terms of U. Pursuing our self-imposed task of compiling a definitive set of tables of symmetric functions we here present the MH-HM tables. In a further publication we shall complete the set by presenting the US-SU tables, the final tables HS-SH being equivalent to the US-SU tables with suitable changes of sign.

2. Many methods can be used for compiling and checking the present tables. One of the simplest is to start from the relationship

$$n!\,h_n = \sum \frac{n!}{1^{\pi_1}\,2^{\pi_2}\dots k^{\pi_k}\,\pi_1!\,\pi_2!\dots\pi_k!}\,s_1^{\pi_1}\,s_2^{\pi_2}\dots s_k^{\pi_k}.$$

The S-functions can be turned into monomial symmetric functions using Part I of our tables and the result follows. Conversely, we have

$$s_n = \sum(-1)^{(\pi_1+\pi_2+\dots+\pi_k+1)}\,\frac{(\pi_1+\pi_2+\dots+\pi_k-1)!\,n}{\pi_1!\,\pi_2!\dots\pi_k!}\,h_1^{\pi_1}\,h_2^{\pi_2}\dots h_k^{\pi_k}.$$

We take the monomial symmetric function and express it in terms of the S-functions from Part I tables. Substitution for s in terms of h gives the required result. The whole of the tables HM was checked by using MacMahon's D operator to build up each table successively. The converse tables MH were checked in the same way, first by using the operator and then by solving the simple equations of relationship which result. For the MH tables we have the additional check that the sum of the coefficients is zero.

3. In the introduction to Parts II and III we mentioned the uses of the U-functions in problems of distribution. The H-functions can be used in an analogous way and we quote therefore only one example. A distribution theorem due to MacMahon can be stated as follows:

'The number of ways of distributing objects of specification $(p_1p_2\cdots p_r)$ into boxes of specification (1^m), no box being empty, is equal to the coefficient of

$$x^{p_1+p_2+\dots+p_r}(p_1p_2\cdots p_r)$$

in the development of the function

$$(h_1x+h_2x^2+\dots)^m.\text{'}$$

When we come to use this theorem it is found that the easiest method of attack is the expression of the powers and products of the H-functions in terms of the M-functions. This is one of the simplest theorems in which the tables will be found of use.

4. In order to save space, when the tables have been set out use has been made of Meyer Hirsch's law of symmetry which states that the coefficient of the monomial symmetric function $(r_1r_2\dots r_t)$ in the expansion of $h_{l_1}h_{l_2}\cdots h_{l_m}$ is equal to the coefficient of the monomial

symmetric function $(l_1 l_2 \dots l_m)$ in the expansion of $h_{r_1} h_{r_2} \dots h_{r_t}$. This enables each table to be written as a triangle and the only difficulty in putting the two parts together has been at the diagonal. In order to express the M-functions in terms of the H-functions we read horizontally until the diagonal in bold type is reached and then continue vertically downwards to the edge of the table. Thus for $w = 5$ we have

$$(31^2) = 3h_1^5 - 13h_2h_1^3 + 10h_2^2h_1 + 12h_3h_1^2 - 8h_3h_2 - 9h_4h_1 + 5h_5.$$

In order to express H-functions in terms of the M-functions we read vertically until the diagonal in bold type is reached and then continue horizontally to the right to the edge of the table. Thus, again for $w = 5$,

$$h_3h_1^2 = 20(1^5) + 13(21^3) + 8(2^21) + 7(31^2) + 4(32) + 3(41) + (5).$$

The two diagonal figures are separated by a bold horizontal rule.

REFERENCES

DAVID, F. N. & KENDALL, M. G. (1949). *Biometrika*, **36**, 431.

DAVID, F. N. & KENDALL, M. G. (1951). *Biometrika*, **38**, 435.

Table 4.2

$w=2$	h_1^2	h_2
(1^2)	**2** **1**	1 ·
(2)	· −1	**1** **2**

Table 4.3

$w=3$	h_1^3	h_2h_1	h_3
(1^3)	**6** **1**	3	1
(21)	· −2	**2** **5**	1
(3)	1	−3	**1** **3**

Table 4.4

$w=4$	h_1^4	$h_2h_1^2$	h_2^2	h_3h_1	h_4
(1^4)	**24** **1**	12	6	4	1
(21^2)	−3	**7** **10**	4	3	1
(2^2)	1	−4	**3** **3**	2	1
(31)	2	−7	2	**2** **7**	1
(4)	−1	4	−2	−4	**1** **4**

Table 4.5

$w=5$	h_1^5	$h_2h_1^3$	$h_2^2h_1$	$h_3h_1^2$	h_3h_2	h_4h_1	h_5
(1^5)	**120** **1**	60	30	20	10	5	1
(21^3)	−4	**33** **17**	18	13	7	4	1
(2^21)	3	−14	**11** **14**	8	5	3	1
(31^2)	3	−13	10	**7** **12**	4	3	1
(32)	−2	10	−11	−8	**3** **11**	2	1
(41)	−2	9	−7	−9	5	**2** **9**	1
(5)	1	−5	5	5	−5	−5	**1** **5**

Table 4.6

$w=6$	h_1^6	$h_2h_1^4$	$h_2^2h_1^2$	h_2^3	$h_3h_1^3$	$h_3h_2h_1$	$h_4h_1^2$	h_3^2	h_4h_2	h_5h_1	h_6
(1^6)	$\frac{720}{1}$	360	180	90	120	60	30	20	15	6	1
(21^4)	−5	$\frac{192}{26}$	102	54	72	38	21	14	11	5	1
(2^21^2)	6	−33	$\frac{58}{46}$	33	42	24	14	10	8	4	1
(2^3)	−1	6	−10	$\frac{21}{4}$	24	15	9	7	6	3	1
(31^3)	4	−21	26	−4	$\frac{34}{19}$	19	13	8	7	4	1
(321)	−6	34	−48	8	−30	$\frac{12}{61}$	8	6	5	3	1
(41^2)	−3	16	−20	4	−15	21	$\frac{7}{15}$	4	4	3	1
(3^2)	1	−6	9	−1	6	−15	−3	$\frac{4}{6}$	3	2	1
(42)	2	−12	19	−6	10	−20	−10	3	$\frac{3}{14}$	2	1
(51)	2	−11	14	−2	11	−17	−11	3	6	$\frac{2}{11}$	1
(6)	−1	6	−9	2	−6	12	6	−3	−6	−6	$\frac{1}{6}$

Table 4.7

$w=7$	h_1^7	$h_2h_1^5$	$h_2^2h_1^3$	$h_2^3h_1$	$h_3h_1^4$	$h_3h_2h_1^2$	$h_3h_2^2$	$h_3^2h_1$	$h_4h_1^3$	$h_4h_2h_1$	h_4h_3	$h_5h_1^2$	h_5h_2	h_6h_1	h_7
(1^7)	$\frac{5040}{1}$	2520	1260	630	840	420	210	140	210	105	35	42	21	7	1
(21^5)	−6	$\frac{1320}{37}$	690	360	480	250	130	90	135	70	25	31	16	6	1
(2^21^3)	10	−64	$\frac{378}{117}$	207	270	148	81	58	84	46	18	22	12	5	1
(2^31)	−4	27	−54	$\frac{120}{30}$	150	87	51	37	51	30	13	15	9	4	1
(31^4)	5	−31	53	−21	$\frac{208}{28}$	114	62	46	73	39	15	21	11	5	1
(321^2)	−12	78	−143	62	−70	$\frac{67}{195}$	39	30	43	25	11	14	8	4	1
(32^2)	3	−21	44	−26	17	−54	$\frac{25}{26}$	19	25	16	8	9	6	3	1
(3^21)	3	−20	37	−14	20	−60	13	$\frac{16}{25}$	20	13	7	8	5	3	1
(41^3)	−4	25	−43	18	−23	55	−14	−14	$\frac{34}{22}$	19	8	13	7	4	1
(421)	6	−40	76	−36	36	−101	30	29	−33	$\frac{12}{64}$	6	8	5	3	1
(43)	−2	14	−28	13	−14	45	−13	−19	11	−26	$\frac{4}{19}$	4	3	2	1
(51^2)	3	−19	33	−14	18	−43	12	10	−18	25	−7	$\frac{7}{18}$	4	3	1
(52)	−2	14	−29	17	−12	36	−17	−7	12	−24	7	−12	$\frac{3}{17}$	2	1
(61)	−2	13	−23	9	−13	33	−7	−10	13	−20	7	−13	7	$\frac{2}{13}$	1
(7)	1	−7	14	−7	7	−21	7	7	−7	14	−7	7	−7	−7	$\frac{1}{7}$

Table 4.8

$w=8$ (i)	h_1^8	$h_2h_1^6$	$h_2^2h_1^4$	$h_2^3h_1^2$	h_2^4	$h_3h_1^5$	$h_3h_2h_1^3$	$h_3h_2^2h_1$	$h_3^2h_1^2$	$h_3^2h_2$
(1^8)	**40320** **1**	20160	10080	5040	2520	6720	3360	1680	1120	560
(21^6)	−7	**10440** **50**	5400	2790	1440	3720	1920	990	680	350
(2^21^4)	15	−110	**2892** **251**	1548	828	2040	1092	584	412	220
(2^31^2)	−10	76	−183	**861** **146**	480	1110	618	345	248	139
(2^4)	1	−8	21	−20	**282** **5**	600	348	204	148	88
(31^5)	6	−43	94	−63	6	**1520** **39**	820	440	320	170
(321^3)	−20	148	−338	240	−24	−134	**463** **489**	260	194	108
(32^21)	12	−93	228	−182	20	81	−322	**154** **258**	116	69
(3^21^2)	6	−45	103	−70	6	44	−165	102	**92** **66**	54
(3^22)	−3	24	−61	50	−4	−22	94	−84	−33	**35** **35**
(41^4)	−5	36	−79	54	−6	−33	111	−66	−35	16
(421^2)	12	−90	209	−154	20	82	−298	192	99	−48
(42^2)	−3	24	−62	56	−12	−20	80	−64	−23	14
(431)	−6	46	−108	76	−8	−46	176	−107	−74	31
(4^2)	1	−8	20	−16	3	8	−32	20	14	−4
(51^3)	4	−29	64	−44	4	27	−91	58	27	−16
(521)	−6	46	−110	84	−8	−42	159	−119	−49	37
(53)	2	−16	40	−31	2	16	−67	54	27	−23
(61^2)	−3	22	−49	34	−4	−21	71	−43	−22	11
(62)	2	−16	41	−36	6	14	−56	48	15	−14
(71)	2	−15	34	−23	2	15	−53	31	19	−8
(8)	−1	8	−20	16	−2	−8	32	−24	−12	8

$w=8$ (ii)	$h_4h_1^4$	$h_4h_2h_1^2$	$h_4h_2^2$	$h_4h_3h_1$	h_4^2	$h_5h_1^3$	$h_5h_2h_1$	h_5h_3	$h_6h_1^2$	h_6h_2	h_7h_1	h_8
(1^8)	1680	840	420	280	70	336	168	56	56	28	8	1
(21^6)	1020	525	270	185	50	228	117	41	43	22	7	1
(2^21^4)	606	324	173	122	36	150	80	30	32	17	6	1
(2^31^2)	354	198	111	80	26	96	54	22	23	13	5	1
(2^4)	204	120	72	52	19	60	36	16	16	10	4	1
(31^5)	500	265	140	100	30	136	71	26	31	16	6	1
(321^3)	286	160	89	66	22	85	47	19	22	12	5	1
(32^21)	162	96	57	43	16	52	31	14	15	9	4	1
(3^21^2)	126	76	45	36	14	44	26	12	14	8	4	1
(3^22)	70	45	29	23	10	26	17	9	9	6	3	1
(41^4)	**209** **31**	115	63	47	16	73	39	15	21	11	5	1
(421^2)	−76	**68** **211**	40	31	12	43	25	11	14	8	4	1
(42^2)	20	−66	**26** **36**	20	9	25	16	8	9	6	3	1
(431)	40	−121	32	**17** **98**	8	20	13	7	8	5	3	1
(4^2)	−8	28	−12	−24	**5** **10**	8	6	4	4	3	2	1
(51^3)	−26	60	−14	−28	4	**34** **26**	19	8	13	7	4	1
(521)	39	−105	28	54	−8	−39	**12** **76**	6	8	5	3	1
(53)	−13	39	−8	−31	4	13	−31	**4** **23**	4	3	2	1
(61^2)	21	−50	14	23	−4	−21	29	−8	**7** **21**	4	3	1
(62)	−14	42	−20	−16	4	14	−28	8	−14	**3** **20**	2	1
(71)	−15	38	−8	−23	4	15	−23	8	−15	8	**2** **15**	1
(8)	8	−24	8	16	−4	−8	16	−8	8	−8	−8	**1** **8**

Table 4.9

$w=9$ (i)	h_1^9	$h_2h_1^7$	$h_2^2h_1^5$	$h_2^3h_1^3$	$h_2^4h_1$	$h_3h_1^6$	$h_3h_2h_1^4$	$h_3h_2^2h_1^2$	$h_3h_2^3$	$h_3^2h_1^3$	$h_3^2h_2h_1$	h_3^3	$h_4h_1^5$
(1^9)	362880 1	181440	90720	45360	22680	60480	30240	15120	7560	10080	5040	1680	15120
(21^7)	−8	93240 65	47880	24570	12600	32760	16800	8610	4410	5880	3010	1050	8820
(2^21^5)	21	−174	25260 478	13320	7020	17640	9300	4900	2580	3420	1800	660	5070
(2^31^3)	−20	170	−484	7227 517	3924	9450	5130	2787	1515	1980	1077	417	2880
(2^41)	5	−44	132	−154	2202 55	5040	2820	1584	894	1140	644	264	1620
(31^6)	7	−57	152	−146	36	12840 52	6840	3630	1920	2600	1370	510	4020
(321^4)	−30	250	−687	686	−177	−228	3764 1039	2064	1128	1508	822	324	2250
(32^21^2)	30	−258	741	−789	222	231	−1110	1173 1297	666	868	493	207	1250
(32^3)	−4	36	−111	134	−50	−30	153	−202	396 50	496	295	132	690
(3^21^3)	10	−84	231	−226	54	81	−374	392	−48	664 149	378	162	950
(3^221)	−12	105	−306	325	−82	−99	494	−597	78	−192	227 327	105	520
(3^3)	1	−9	27	−29	6	9	−48	63	−6	21	−45	55 10	210
(41^5)	−6	49	−131	127	−33	−45	195	−195	27	−68	78	−6	1545 42
(421^3)	20	−168	466	−474	132	154	−699	740	−112	249	−304	24	−143
(42^21)	−12	105	−309	344	−110	−93	451	−536	98	−154	222	−18	87
(431^2)	−12	102	−284	283	−74	−100	463	−477	66	−189	216	−18	91
(432)	6	−54	164	−188	60	50	−256	324	−60	96	−157	15	−44
(4^21)	3	−26	74	−76	22	26	−122	124	−19	52	−51	3	−26
(51^4)	5	−41	110	−107	27	38	−165	169	−23	56	−70	6	−36
(521^2)	−12	102	−287	297	−82	−94	434	−479	74	−151	204	−18	88
(52^2)	3	−27	83	−100	38	23	−115	148	−38	35	−51	3	−23
(531)	6	−52	148	−150	36	52	−250	275	−34	104	−142	15	−46
(54)	−2	18	−54	60	−19	−18	90	−104	19	−38	47	−3	18
(61^3)	−4	33	−89	87	−22	−31	135	−139	18	−46	60	−6	30
(621)	6	−52	150	−160	44	48	−229	268	−38	79	−127	15	−45
(63)	−2	18	−54	59	−15	−18	93	−117	15	−39	72	−12	15
(71^2)	3	−25	68	−67	18	24	−105	106	−16	37	−42	3	−24
(72)	−2	18	−55	65	−23	−16	80	−103	23	−25	41	−3	16
(81)	−2	17	−47	46	−11	−17	77	−78	9	−30	35	−3	17
(9)	1	−9	27	−30	9	9	−45	54	−9	18	−27	3	−9

$w=9$ (ii)	$h_4h_2h_1^3$	$h_4h_2^2h_1$	$h_4h_3h_1^2$	$h_4h_3h_2$	$h_4^2h_1$	$h_5h_1^4$	$h_5h_2h_1^2$	$h_5h_2^2$	$h_5h_3h_1$	h_5h_4	$h_6h_1^3$	$h_6h_2h_1$	h_6h_3	$h_7h_1^2$	h_7h_2	h_8h_1	h_9
(1^9)	7560	3780	2520	1260	630	3024	1512	756	504	126	504	252	84	72	36	9	1
(21^7)	4515	2310	1575	805	420	1932	987	504	343	91	357	182	63	57	29	8	1
(2^21^5)	2670	1405	980	515	280	1206	634	333	232	66	246	129	47	44	23	7	1
(2^31^3)	1566	852	606	330	186	738	402	219	156	48	165	90	35	33	18	6	1
(2^41)	912	516	372	212	123	444	252	144	104	35	108	62	26	24	14	5	1
(31^6)	2115	1110	785	410	230	1044	543	282	197	56	229	118	42	43	22	7	1
(321^4)	1229	669	486	263	154	626	339	183	132	41	151	81	31	32	17	6	1
(32^21^2)	710	403	298	169	102	370	210	119	88	30	97	55	23	23	13	5	1
(32^3)	408	243	181	109	67	216	129	78	58	22	61	37	17	16	10	4	1
(3^21^3)	548	313	240	135	86	302	172	97	74	26	86	48	20	22	12	5	1
(3^221)	313	188	145	87	56	174	105	63	49	19	53	32	15	15	9	4	1
(3^3)	135	87	69	45	30	78	51	33	27	12	27	18	10	9	6	3	1
(41^5)	840	455	335	180	110	501	266	141	101	31	136	71	26	31	16	6	1
(421^3)	479 520	272	206	116	74	287	161	90	67	23	85	47	19	22	12	5	1
(42^21)	−344	163 270	125	75	49	163	97	58	44	17	52	31	14	15	9	4	1
(431^2)	−344	217	101 277	60	42	127	77	46	37	15	44	26	12	14	8	4	1
(432)	184	−154	−131	39 112	27	71	46	30	24	11	26	17	9	9	6	3	1
(4^21)	104	−69	−91	38	21 39	48	32	21	18	9	20	13	7	8	5	3	1
(51^4)	119	−71	−71	36	18	209 35	115	63	47	16	73	39	15	21	11	5	1
(521^2)	−317	209	200	−113	−55	−84	68 228	40	31	12	43	25	11	14	8	4	1
(52^2)	94	−80	−54	42	19	21	−68	26 38	20	9	25	16	8	9	6	3	1
(531)	175	−110	−143	74	42	42	−123	28	17 97	8	20	13	7	8	5	3	1
(54)	−76	57	67	−38	−29	−14	47	−19	−38	5 29	8	6	4	4	3	2	1
(61^3)	−99	61	56	−31	−13	−30	69	−16	−32	9	34 30	19	8	13	7	4	1
(621)	166	−120	−95	68	22	45	−121	32	62	−18	−45	12 88	6	8	5	3	1
(63)	−60	45	45	−36	−9	−15	45	−9	−36	9	15	−36	4 27	4	3	2	1
(71^2)	81	−49	−50	25	13	24	−57	16	26	−9	−24	33	−9	7 24	4	3	1
(72)	−64	55	34	−32	−9	−16	48	−23	−18	9	16	−32	9	−16	3 23	2	1
(81)	−60	35	43	−18	−13	−17	43	−9	−26	9	17	−26	9	−17	9	2 17	1
(9)	36	−27	−27	18	9	9	−27	9	18	−9	−9	18	−9	9	−9	−9	1 9

Table 4.10

$w=10$ (i)	h_1^{10}	$h_2h_1^8$	$h_2^2h_1^6$	$h_2^3h_1^4$	$h_2^4h_1^2$	h_2^5	$h_3h_1^7$	$h_3h_2h_1^5$	$h_3h_2^2h_1^3$
(1^{10})	**3628800** 1	1814400	907200	453600	226800	113400	604800	302400	151200
(21^8)	−9	**927360** 82	473760	241920	123480	63000	322560	164640	84000
(2^21^6)	28	−259	**247320** 834	129060	67320	35100	171360	89400	46620
(2^31^4)	−35	330	−1090	**68868** 1476	36756	19620	90720	48420	25848
(2^41^2)	15	−145	496	−708	**20100** 371	11010	47880	26160	14316
(2^5)	−1	10	−36	56	−35	**6210** 6	25200	14100	7920
(31^7)	8	−73	230	−290	124	−8	**122640** 67	64680	34020
(321^5)	−42	390	−1256	1628	−720	48	−358	**34960** 1965	18850
(32^21^3)	60	−570	1892	−2555	1194	−84	518	−2946	**10434** 4655
(32^31)	−20	196	−680	982	−512	40	−172	1023	−1746
(3^21^4)	15	−140	451	−578	246	−15	134	−741	1102
(3^221^2)	−30	288	−963	1296	−582	36	−273	1584	−2531
(3^22^2)	6	−60	213	−315	166	−10	54	−333	596
(3^31)	4	−39	132	−177	74	−4	39	−234	383
(41^6)	−7	64	−202	256	−112	8	−59	313	−449
(421^4)	30	−280	907	−1188	543	−42	258	−1412	2106
(42^21^2)	−30	288	−969	1338	−660	60	−261	1491	−2368
(42^3)	4	−40	143	−218	130	−20	34	−204	352
(431^3)	−20	188	−610	790	−348	24	−182	1006	−1485
(4321)	24	−234	798	−1106	528	−40	222	−1306	2124
(43^2)	−3	30	−105	148	−68	4	−30	186	−320
(4^21^2)	6	−57	187	−246	114	−10	56	−310	452
(4^22)	−3	30	−106	156	−86	12	−28	168	−280
(51^5)	6	−55	174	−221	96	−6	51	−271	393
(521^3)	−20	188	−614	812	−372	24	−174	961	−1463
(52^21)	12	−117	402	−572	292	−20	105	−613	1020
(531^2)	12	−114	374	−489	214	−12	112	−629	950
(532)	−6	60	−212	310	−160	8	−56	342	−600
(541)	−6	58	−194	260	−120	8	−58	330	−500
(5^2)	1	−10	35	−50	25	−1	10	−60	100
(61^4)	−5	46	−146	186	−81	6	−43	229	−333
(621^2)	12	−114	377	−506	236	−20	106	−594	920
(62^2)	−3	30	−107	162	−94	12	−26	156	−272
(631)	−6	58	−194	258	−112	8	−58	336	−525
(64)	2	−20	70	−100	51	−6	20	−120	196
(71^3)	4	−37	118	−151	66	−4	35	−187	273
(721)	−6	58	−196	270	−128	8	−54	311	−503
(73)	2	−20	70	−99	46	−2	20	−123	212
(81^2)	−3	28	−90	116	−52	4	−27	145	−210
(82)	2	−20	71	−106	59	−6	18	−108	188
(91)	2	−19	62	−80	34	−2	19	−105	154
(10)	−1	10	−35	50	−25	2	−10	60	−100

Table 4.10 (*continued*)

$w = 10$ (ii)	$h_3h_2^3h_1$	$h_3^2h_1^4$	$h_3^2h_2h_1^2$	$h_3^2h_2^2$	$h_3^3h_1$	$h_4h_1^6$	$h_4h_2h_1^4$	$h_4h_2^2h_1^2$	$h_4h_2^3$	$h_4h_3h_1^3$
(1^{10})	75600	100800	50400	25200	16800	151200	75600	37800	18900	25200
(21^8)	42840	57120	29120	14840	10080	85680	43680	22260	11340	15120
(2^21^6)	24300	32280	16820	8760	6060	48060	25050	13050	6795	9030
(2^31^4)	13800	18180	9708	5184	3648	26730	14274	7623	4071	5364
(2^41^2)	7848	10200	5596	3076	2196	14760	8088	4440	2442	3168
(2^5)	4470	5700	3220	1830	1320	8100	4560	2580	1470	1860
(31^7)	17850	24080	12600	6580	4620	36960	19320	10080	5250	7070
(321^5)	10140	13560	7280	3900	2790	20340	10930	5860	3135	4195
(32^21^3)	5769	7600	4200	2318	1686	11130	6154	3399	1875	2472
(32^31)	$\frac{3288}{774}$	4240	2418	1382	1017	6060	3450	1968	1125	1447
(3^21^4)	−363	$\frac{5672}{299}$	3156	1748	1296	8300	4650	2590	1435	1932
(3^221^2)	902	−637	$\frac{1818}{1527}$	1044	786	4490	2594	1495	859	1126
(3^22^2)	−282	123	−330	$\frac{626}{120}$	474	2420	1442	862	517	652
(3^31)	−122	104	−273	48	$\frac{370}{65}$	1770	1074	651	393	504
(41^6)	151	−116	228	−45	−30	$\frac{13290}{55}$	7155	3840	2055	2795
(421^4)	−744	529	−1090	227	148	−240	$\frac{3985}{1091}$	2212	1224	1645
(42^21^2)	910	−551	1236	−282	−174	243	−1166	$\frac{1273}{1366}$	732	960
(42^3)	−160	69	−164	42	20	−34	178	−248	$\frac{441}{76}$	556
(431^3)	502	−410	829	−162	−122	168	−777	814	−112	$\frac{749}{622}$
(4321)	−796	524	−1248	276	210	−198	978	−1164	184	−752
(43^2)	116	−83	228	−48	−55	24	−124	156	−20	104
(4^21^2)	−156	129	−240	45	30	−56	266	−284	47	−224
(4^22)	112	−66	152	−29	−22	28	−148	200	−54	112
(51^5)	−135	99	−200	43	26	−48	206	−205	26	−140
(521^3)	542	−355	755	−182	−98	163	−735	774	−104	514
(52^21)	−446	217	−510	162	60	−99	476	−554	84	−323
(531^2)	−332	259	−543	120	78	−103	477	−493	58	−389
(532)	280	−130	331	−122	−41	50	−250	298	−36	189
(541)	181	−140	277	−61	−35	58	−280	299	−40	245
(5^2)	−45	25	−55	20	5	−10	50	−55	5	−45
(61^4)	110	−84	175	−33	−26	41	−176	180	−26	115
(621^2)	−328	221	−498	102	78	−100	459	−512	86	−307
(62^2)	128	−52	128	−40	−16	26	−132	176	−48	80
(631)	169	−144	336	−57	−65	52	−249	279	−40	201
(64)	−72	52	−120	21	22	−20	104	−132	30	−88
(71^3)	−94	69	−142	30	20	−34	146	−147	18	−97
(721)	193	−115	277	−67	−44	51	−239	273	−36	158
(73)	−82	53	−144	36	31	−17	85	−102	10	−68
(81^2)	73	−55	107	−23	−13	27	−118	119	−18	83
(82)	−88	37	−92	31	10	−18	90	−116	26	−56
(91)	−49	43	−87	15	13	−19	86	−87	10	−67
(10)	40	−25	60	−15	−10	10	−50	60	−10	40

Table 4.10 (*continued*)

$w = 10$ (iii)	$h_4h_3h_2h_1$	$h_4h_3^2$	$h_4^2h_1^2$	$h_4^2h_2$	$h_5h_1^5$	$h_5h_2h_1^3$	$h_5h_2^2h_1$	$h_5h_3h_1^2$	$h_5h_3h_2$	$h_5h_4h_1$	h_5^2
(1^{10})	12600	4200	6300	3150	30240	15120	7560	5040	2520	1260	252
(21^8)	7700	2660	3990	2030	18480	9408	4788	3248	1652	854	182
(2^21^6)	4700	1690	2520	1310	11100	5778	3006	2078	1080	578	132
(2^31^4)	2865	1077	1584	846	6570	3510	1875	1320	705	390	96
(2^41^2)	1744	688	990	547	3840	2112	1164	832	460	262	70
(2^5)	1060	440	615	355	2220	1260	720	520	300	175	51
(31^7)	3675	1330	2030	1050	9240	4788	2478	1722	889	483	112
(321^5)	2240	850	1280	680	5380	2872	1530	1089	578	327	82
(32^21^3)	1363	545	800	440	3100	1710	942	684	376	220	60
(32^31)	828	350	496	285	1770	1011	579	426	245	147	44
(3^21^4)	1066	432	652	356	2460	1366	754	560	306	186	52
(3^221^2)	647	279	402	230	1390	802	461	348	199	124	38
(3^22^2)	392	180	246	149	780	468	282	214	130	82	28
(3^31)	306	145	198	120	600	366	222	174	105	69	24
(41^6)	1490	570	890	470	4050	2139	1128	803	422	242	62
(421^4)	907	367	560	306	2275	1249	684	501	273	164	46
(42^21^2)	551	237	348	199	1270	726	415	310	177	110	34
(42^3)	334	153	214	130	705	420	252	190	115	73	25
(431^3)	431	189	286	162	970	564	325	252	143	94	30
(4321)	$\frac{261}{1157}$	123	174	105	535	325	197	154	93	62	22
(43^2)	−195	$\frac{65}{55}$	84	55	220	143	93	75	49	34	14
(4^21^2)	249	−27	$\frac{126}{98}$	75	350	218	134	110	66	48	18
(4^22)	−180	22	−49	$\frac{49}{54}$	190	124	81	66	43	31	13
(51^5)	166	−20	44	−20	$\frac{1546}{46}$	841	456	336	181	111	32
(521^3)	−648	80	−168	80	−155	$\frac{480}{561}$	273	207	117	75	24
(52^21)	465	−57	109	−58	95	−379	$\frac{164}{311}$	126	76	50	18
(531^2)	470	−63	136	−58	95	−358	232	$\frac{102}{282}$	61	43	16
(532)	−295	41	−59	28	−50	215	−193	−154	$\frac{40}{157}$	28	12
(541)	−280	32	−108	44	−50	201	−141	−172	90	$\frac{22}{143}$	10
(5^2)	55	−5	20	−5	10	−45	40	40	−35	−35	$\frac{6}{15}$
(61^4)	−146	20	−34	20	−40	130	−75	−75	35	35	−5
(621^2)	431	−63	95	−66	96	−339	215	205	−105	−102	15
(62^2)	−128	16	−31	30	−24	96	−80	−50	36	32	−5
(631)	−291	53	−63	44	−48	177	−102	−141	61	73	−10
(64)	132	−22	39	−34	16	−64	42	54	−20	−44	5
(71^3)	122	−17	28	−14	34	−112	69	63	−35	−29	5
(721)	−247	41	−43	28	−51	188	−136	−107	77	49	−10
(73)	123	−31	15	−10	17	−68	51	51	−41	−20	5
(81^2)	−94	10	−28	14	−27	91	−55	−56	28	29	−5
(82)	92	−10	19	−18	18	−72	62	38	−36	−20	5
(91)	78	−10	24	−10	19	−67	39	48	−20	−29	5
(10)	−60	10	−15	10	−10	40	−30	−30	20	20	−5

Table 4.10 (*continued*)

$w = 10$ (iv)	$h_6h_1^4$	$h_6h_2h_1^2$	$h_6h_2^2$	$h_6h_3h_1$	h_6h_4	$h_7h_1^3$	$h_7h_2h_1$	h_7h_3	$h_8h_1^2$	h_8h_2	h_9h_1	h_{10}
(1^{10})	5040	2520	1260	840	210	720	360	120	90	45	10	1
(21^8)	3360	1708	868	588	154	528	268	92	73	37	9	1
(2^21^6)	2190	1138	591	408	113	378	196	70	58	30	8	1
(2^31^4)	1398	747	399	281	83	264	141	53	45	24	7	1
(2^41^2)	876	484	268	192	61	180	100	40	34	19	6	1
(2^5)	540	310	180	130	45	120	70	30	25	15	5	1
(31^7)	1960	1008	518	357	98	358	183	64	57	29	8	1
(321^5)	1230	652	345	244	72	247	130	48	44	23	7	1
(32^21^3)	758	417	229	166	53	166	91	36	33	18	6	1
(32^31)	460	264	152	112	39	109	63	27	24	14	5	1
(3^21^4)	646	354	193	142	46	152	82	32	32	17	6	1
(3^221^2)	386	222	127	96	34	98	56	24	23	13	5	1
(3^22^2)	228	138	84	64	25	62	38	18	16	10	4	1
(3^31)	186	114	69	55	22	54	33	16	15	9	4	1
(41^6)	1045	544	283	198	57	229	118	42	43	22	7	1
(421^4)	627	340	184	133	42	151	81	31	32	17	6	1
(42^21^2)	371	211	120	89	31	97	55	23	23	13	5	1
(42^3)	217	130	79	59	23	61	37	17	16	10	4	1
(431^3)	303	173	98	75	27	86	48	20	22	12	5	1
(4321)	175	106	64	50	20	53	32	15	15	9	4	1
(43^2)	79	52	34	28	13	27	18	10	9	6	3	1
(4^21^2)	128	78	47	38	16	44	26	12	14	8	4	1
(4^22)	72	47	31	25	12	26	17	9	9	6	3	1
(51^5)	501	266	141	101	31	136	71	26	31	16	6	1
(521^3)	287	161	90	67	23	85	47	19	22	12	5	1
(52^21)	163	97	58	44	17	52	31	14	15	9	4	1
(531^2)	127	77	46	37	15	44	26	12	14	8	4	1
(532)	71	46	30	24	11	26	17	9	9	6	3	1
(541)	48	32	21	18	9	20	13	7	8	5	3	1
(5^2)	16	12	9	8	5	8	6	4	4	3	2	1
(61^4)	$\frac{209}{40}$	115	63	47	16	73	39	15	21	11	5	1
(621^2)	−96	$\frac{68}{261}$	40	31	12	43	25	11	14	8	4	1
(62^2)	24	−78	$\frac{26}{44}$	20	9	25	16	8	9	6	3	1
(631)	48	−141	32	$\frac{17}{112}$	8	20	13	7	8	5	3	1
(64)	−16	54	−22	−44	$\frac{5}{34}$	8	6	4	4	3	2	1
(71^3)	−34	78	−18	−36	10	$\frac{34}{34}$	19	8	13	7	4	1
(721)	51	−137	36	70	−20	−51	$\frac{12}{100}$	6	8	5	3	1
(73)	−17	51	−10	−41	10	17	−41	$\frac{4}{31}$	4	3	2	1
(81^2)	27	−64	18	29	−10	−27	37	−10	$\frac{7}{27}$	4	3	1
(82)	−18	54	−26	−20	10	18	−36	10	−18	$\frac{3}{26}$	2	1
(91)	−19	48	−10	−29	10	19	−29	10	−19	10	$\frac{2}{19}$	1
(10)	10	−30	10	20	−10	−10	20	−10	10	−10	−10	$\frac{1}{10}$

Table 4.11

$w = 11$ (i)	h_1^{11}	$h_2h_1^9$	$h_2^2h_1^7$	$h_2^3h_1^5$	$h_2^4h_1^3$	$h_2^5h_1$	$h_3h_1^8$	$h_3h_2h_1^6$	$h_3h_2^2h_1^4$	$h_3h_2^3h_1^2$
(1^{11})	39916800 1	19958400	9979200	4989600	2494800	1247400	6652800	3326400	1663200	831600
(21^9)	−10	10160640 101	5171040	2630880	1338120	680400	3507840	1784160	907200	461160
(2^21^7)	36	−368	2678760 1361	1387260	718200	371700	1844640	955080	494340	255780
(2^31^5)	−56	581	−2190	731520 3614	385740	203400	967680	510300	269100	141900
(2^41^3)	35	−370	1430	−2444	207324 1742	111510	506520	272160	146340	78744
(2^51)	−6	65	−260	468	−364	61260 91	264600	144900	79500	43710
(31^8)	9	−91	331	−519	325	−55	1303680 84	682080	356160	185640
(321^6)	−56	574	−2123	3398	−2182	380	−530	363840 3411	193740	102990
(32^21^4)	105	−1095	4140	−6818	4544	−828	1005	−6634	105284 13359	57156
(32^31^2)	−60	640	−2494	4284	−3035	602	−576	3926	−8298	31731 5583
(32^4)	5	−55	224	−412	329	−85	47	−332	744	−562
(3^21^5)	21	−216	799	−1270	798	−132	206	−1332	2580	−1488
(3^221^3)	−60	630	−2392	3929	−2566	438	−598	3994	−8085	4926
(3^22^21)	30	−324	1278	−2214	1557	−282	300	−2091	4527	−3095
(3^31^2)	10	−106	405	−663	420	−66	105	−714	1458	−856
(3^32)	−4	44	−177	312	−219	34	−42	303	−684	487
(41^7)	−8	81	−295	464	−294	52	−75	471	−888	510
(421^5)	42	−432	1604	−2584	1688	−312	400	−2569	4980	−2961
(42^21^3)	−60	630	−2402	4007	−2745	546	−578	3824	−7717	4845
(42^31)	20	−216	856	−1510	1128	−260	192	−1319	2820	−1954
(431^4)	−30	310	−1152	1843	−1178	207	−298	1925	−3713	2151
(4321^2)	60	−636	2442	−4074	2742	−516	606	−4074	8296	−5118
(432^2)	−12	132	−534	963	−734	170	−120	849	−1879	1352
(43^21)	−12	129	−501	837	−547	94	−129	891	−1849	1103
(4^21^3)	10	−104	389	−628	411	−78	101	−652	1250	−729
(4^221)	−12	129	−504	862	−608	130	−123	836	−1722	1088
(4^23)	3	−33	132	−230	162	−34	33	−234	502	−316
(51^6)	7	−71	259	−408	258	−44	66	−415	787	−458
(521^4)	−30	310	−1157	1875	−1229	219	−288	1860	−3647	2220
(52^21^2)	30	−318	1227	−2079	1449	−282	291	−1950	4027	−2645
(52^3)	−4	44	−179	329	−264	70	−38	266	−589	442
(531^3)	20	−208	778	−1252	800	−132	202	−1316	2571	−1526
(5321)	−24	258	−1008	1718	−1178	204	−246	1690	−3568	2354
(53^2)	3	−33	132	−229	155	−22	33	−237	525	−355
(541^2)	−12	126	−476	776	−511	94	−124	810	−1576	943
(542)	6	−66	266	−476	360	−84	62	−434	940	−660
(5^21)	3	−32	123	−204	135	−22	32	−214	430	−270
(61^5)	−6	61	−223	352	−223	39	−57	359	−682	393
(621^3)	20	−208	782	−1278	846	−156	194	−1263	2500	−1522
(62^21)	−12	129	−507	881	−636	130	−117	799	−1699	1164
(631^2)	−12	126	−476	773	−497	86	−124	819	−1619	949
(632)	6	−66	266	−474	348	−68	62	−440	978	−704
(641)	6	−64	246	−408	272	−52	64	−428	852	−505
(65)	−2	22	−88	154	−110	21	−22	154	−330	225
(71^4)	5	−51	187	−296	188	−33	48	−303	577	−333
(721^2)	−12	126	−479	793	−533	102	−118	778	−1563	961
(72^2)	3	−33	134	−245	194	−50	29	−203	449	−332
(731)	6	−64	246	−406	262	−44	64	−434	883	−520
(74)	−2	22	−88	154	−111	25	−22	154	−326	208
(81^3)	−4	41	−151	240	−153	26	−39	247	−472	277
(821)	6	−64	248	−420	288	−52	60	−405	842	−545
(83)	−2	22	−88	153	−105	17	−22	157	−345	230
(91^2)	3	−31	115	−184	119	−22	30	−191	364	−213
(92)	−2	22	−89	161	−124	29	−20	140	−309	227
$(10, 1)$	−2	21	−79	127	−80	13	−21	137	−265	150
(11)	1	−11	44	−77	55	−11	11	−77	165	−110

Table 4.11 (*continued*)

$w = 11$ (ii)	$h_3h_2^4$	$h_3^2h_1^5$	$h_3^2h_2h_1^3$	$h_3^2h_2^2h_1$	$h_3^3h_1^2$	$h_3^3h_2$	$h_4h_1^7$	$h_4h_2h_1^5$	$h_4h_2^2h_1^3$	$h_4h_2^3h_1$
(1^{11})	415800	1108800	554400	277200	184800	92400	1663200	831600	415800	207900
(21^9)	234360	614880	312480	158760	107520	54600	922320	468720	238140	120960
(2^21^7)	132300	340200	175980	91000	62580	32340	507780	262710	135870	70245
(2^31^5)	74820	187740	99000	52200	36420	19200	277830	146520	77265	40740
(2^41^3)	42402	103320	55620	29964	21180	11424	151200	81360	43812	23610
(2^51)	24090	56700	31200	17210	12300	6810	81900	45000	24780	13680
(31^8)	96600	249760	129920	67480	47040	24360	381360	198240	102900	53340
(321^6)	54660	137720	73100	38740	27420	14490	207060	109950	58290	30855
(32^21^4)	30996	75700	41068	22256	15972	8640	111930	60750	32939	17841
(32^31^2)	17622	41480	23032	12794	9288	5163	60270	33450	18573	10317
(32^4)	**10050** **85**	22660	12892	7358	5388	3090	32340	18360	10452	5970
(3^21^5)	117	**55480** **546**	30340	16540	12060	6540	82180	45110	24670	13445
(3^221^3)	−402	−1638	**17010** **5235**	9516	7026	3918	44030	24740	13873	7762
(3^22^21)	272	828	−2872	**5478** **1923**	4080	2352	23520	13530	7788	4485
(3^31^2)	62	315	−1043	540	**3098** **245**	1792	17010	9930	5787	3363
(3^32)	−34	−126	472	−357	−98	**1081** **85**	9030	5400	3237	1944
(41^7)	−44	−182	517	−255	−87	33	**129990** **70**	69405	36960	19635
(421^5)	270	999	−2934	1503	504	−198	−373	**38010** **2044**	20760	11310
(42^21^3)	−486	−1478	4561	−2483	−795	331	539	−3063	**11642** **4843**	6519
(42^31)	240	489	−1598	974	262	−124	−183	1092	−1870	**3762** **840**
(431^4)	−183	−793	2318	−1145	−428	158	277	−1532	2278	−778
(4321^2)	476	1676	−5305	2820	1041	−408	−552	3194	−5096	1896
(432^2)	−170	−327	1126	−732	−198	102	111	−682	1208	−560
(43^21)	−90	−400	1341	−675	−327	111	111	−662	1085	−380
(4^21^3)	68	271	−761	365	131	−44	−100	560	−839	300
(4^221)	−118	−344	1076	−552	−204	66	120	−716	1192	−488
(4^23)	34	106	−364	180	92	−23	−30	188	−332	136
(51^6)	38	159	−457	237	75	−33	−62	327	−467	155
(521^4)	−195	−717	2140	−1177	−354	170	270	−1471	2185	−764
(52^21^2)	262	740	−2348	1425	384	−213	−273	1553	−2457	946
(52^3)	−70	−93	308	−200	−44	22	38	−232	414	−204
(531^3)	120	544	−1616	868	287	−136	−188	1037	−1524	504
(5321)	−196	−686	2276	−1446	−414	258	222	−1290	2058	−752
(53^2)	22	105	−385	267	88	−65	−27	162	−265	88
(541^2)	−86	−340	967	−498	−159	66	124	−698	1041	−365
(542)	84	172	−560	343	90	−43	−62	380	−660	298
(5^21)	21	91	−270	161	42	−26	−32	182	−270	89
(61^5)	−33	−138	402	−202	−70	28	54	−285	412	−140
(621^3)	136	489	−1493	808	263	−118	−183	1006	−1528	556
(62^21)	−118	−298	986	−642	−162	102	111	−644	1060	−444
(631^2)	−78	−347	1062	−540	−210	84	115	−646	980	−337
(632)	68	174	−615	447	107	−85	−56	336	−570	242
(641)	46	186	−557	270	109	−37	−64	372	−583	214
(65)	−21	−66	215	−141	−37	26	22	−132	215	−84
(71^4)	29	117	−342	169	61	−22	−46	243	−352	118
(721^2)	−94	−303	947	−498	−177	66	112	−625	972	−357
(72^2)	50	72	−243	155	39	−17	−29	176	−314	154
(731)	42	190	−616	306	139	−46	−58	335	−528	178
(74)	−25	−68	222	−108	−50	11	22	−136	236	−100
(81^3)	−22	−96	279	−148	−47	22	38	−201	289	−97
(821)	46	157	−515	313	93	−52	−57	324	−515	192
(83)	−17	−69	250	−162	−58	35	19	−114	190	−68
(91^2)	20	76	−215	108	35	−14	−30	161	−233	81
(92)	−29	−51	173	−120	−25	17	20	−120	209	−98
$(10, 1)$	−11	−58	170	−81	−32	11	21	−116	170	−54
(11)	11	33	−110	66	22	−11	−11	66	−110	44

Table 4.11 (*continued*)

$w = 11$ (iii)	$h_4h_3h_1^4$	$h_4h_3h_2h_1^2$	$h_4h_3h_2^2$	$h_4h_3^2h_1$	$h_4^2h_1^3$	$h_4^2h_2h_1$	$h_4^2h_3$	$h_5h_1^6$	$h_5h_2h_1^4$	$h_5h_2^2h_1^2$	$h_5h_2^3$	$h_5h_3h_1^3$
(1^{11})	277200	138600	69300	46200	69300	34650	11550	332640	166320	83160	41580	55440
(21^9)	161280	81900	41580	28140	42210	21420	7350	196560	99792	50652	25704	34272
(2^21^7)	93450	48300	24955	17150	25620	13230	4690	114660	59262	30618	15813	21042
(2^31^5)	53910	28425	14985	10455	15480	8160	3000	66150	34884	18393	9696	12834
(2^41^3)	30960	16692	9006	6372	9306	5025	1923	37800	20376	10992	5934	7776
(2^51)	17700	9780	5420	3880	5565	3090	1235	21420	11820	6540	3630	4680
(31^8)	71680	37100	19180	13300	20230	10430	3710	92400	47712	24612	12684	17024
(321^6)	41270	21815	11515	8120	12230	6440	2380	52620	27798	14664	7725	10334
(32^21^4)	23642	12797	6919	4957	7344	3966	1530	29730	16094	8703	4701	6234
(32^31^2)	13478	7489	4163	3023	4382	2437	985	16680	9264	5148	2862	3736
(32^4)	7648	4372	2510	1840	2599	1495	635	9300	5304	3036	1746	2224
(3^21^5)	18050	9810	5315	3860	5820	3140	1220	23060	12574	6830	3697	4978
(3^221^3)	10254	5733	3197	2359	3458	1926	788	12830	7192	4021	2242	2972
(3^22^21)	5798	3342	1927	1438	2042	1179	509	7100	4094	2362	1363	1762
(3^31^2)	4386	2553	1479	1127	1602	930	410	5370	3138	1827	1059	1392
(3^32)	2460	1482	894	685	936	567	265	2940	1770	1068	645	816
(41^7)	26635	14105	7455	5320	8260	4340	1610	37590	19761	10374	5439	7343
(421^5)	15215	8265	4480	3255	4970	2680	1040	20805	11256	6078	3276	4397
(42^21^3)	8642	4830	2696	1990	2964	1649	673	11460	6386	3555	1977	2616
(42^31)	4883	2815	1626	1214	1754	1012	436	6285	3609	2076	1197	1546
(431^4)	**6573** **1237**	3697	2070	1555	2356	1310	540	8615	4871	2738	1531	2069
(4321^2)	−2551	**2151** **5904**	1248	951	1384	801	351	4700	2742	1595	925	1218
(432^2)	509	−1320	**755** **426**	579	808	489	228	2555	1538	928	562	712
(43^21)	576	−1464	294	**456** **488**	630	385	185	1890	1159	709	432	557
(4^21^3)	−465	920	−188	−188	**950** **200**	549	243	2990	1782	1052	616	834
(4^221)	578	−1360	325	320	−240	**334** **383**	159	1610	992	609	373	484
(4^23)	−166	456	−111	−161	60	−119	**85** **69**	630	410	267	174	216
(51^6)	−239	472	−95	−91	84	−96	22	**13326** **59**	7185	3864	2073	2819
(521^4)	1090	−2249	485	440	−391	470	−110	−256	**4010** **1160**	2232	1239	1665
(52^21^2)	−1135	2532	−624	−495	413	−552	132	259	−1240	**1289** **1447**	744	976
(52^3)	157	−392	134	66	−65	109	−25	−34	175	−238	**450** **70**	568
(531^3)	−845	1708	−340	−360	316	−360	88	176	−815	859	−106	**765** **651**
(5321)	1024	−2363	550	529	−359	484	−130	−214	1060	−1257	166	−826
(53^2)	−140	357	−72	−111	41	−60	23	27	−140	174	−16	121
(541^2)	593	−1151	239	217	−262	288	−61	−112	535	−578	86	−452
(542)	−300	724	−214	−137	131	−210	50	54	−280	362	−84	214
(5^21)	−160	307	−63	−53	72	−68	11	32	−160	181	−21	144
(61^5)	204	−417	85	86	−69	86	−22	−52	221	−219	29	−146
(621^3)	−731	1576	−352	−330	254	−336	88	175	−784	828	−118	533
(62^21)	452	−1060	296	206	−163	235	−59	−107	516	−612	100	−342
(631^2)	529	−1140	231	271	−191	250	−73	−107	492	−510	66	−393
(632)	−255	630	−192	−129	87	−126	34	56	−285	354	−50	216
(641)	−321	684	−138	−159	141	−186	50	56	−269	291	−46	230
(65)	115	−252	63	48	−52	63	−11	−22	115	−141	21	−104
(71^4)	−174	363	−73	−80	57	−74	22	45	−191	192	−27	122
(721^2)	453	−1037	233	244	−150	222	−73	−108	489	−535	88	−317
(72^2)	−117	302	−104	−61	45	−79	25	27	−135	178	−50	76
(731)	−281	675	−132	−201	91	−144	62	54	−253	273	−36	199
(74)	118	−300	75	89	−50	89	−39	−18	90	−108	25	−72
(81^3)	146	−293	59	59	−50	60	−15	−38	163	−164	20	−108
(821)	−233	548	−126	−124	76	−114	34	57	−267	305	−40	176
(83)	95	−252	57	81	−26	45	−23	−19	95	−114	11	−76
(91^2)	−122	237	−51	−43	46	−52	11	30	−131	132	−20	92
(92)	82	−204	69	33	−31	51	−11	−20	100	−129	29	−62
$(10, 1)$	95	−192	33	43	−37	43	−11	−21	95	−96	11	−74
(11)	−55	132	−33	−33	22	−33	11	11	−55	66	−11	44

Table 4.11 (*continued*)

$w = 11$ (iv)	$h_5h_3h_2h_1$	$h_5h_3^2$	$h_5h_4h_1^2$	$h_5h_4h_2$	$h_5^2h_1$	$h_6h_1^5$	$h_6h_2h_1^3$	$h_6h_2^2h_1$	$h_6h_3h_1^2$	$h_6h_3h_2$	$h_6h_4h_1$	h_6h_5
(1^{11})	27720	9240	13860	6930	2772	55440	27720	13860	9240	4620	2310	462
(21^9)	17388	5964	8946	4536	1890	35280	17892	9072	6132	3108	1596	336
(2^21^7)	10864	3850	5754	2968	1288	22050	11382	5873	4032	2079	1099	245
(2^31^5)	6765	2487	3684	1941	876	13560	7149	3768	2629	1385	754	179
(2^41^3)	4200	1608	2346	1269	594	8220	4440	2400	1700	920	515	131
(2^51)	2600	1040	1485	830	401	4920	2730	1520	1090	610	350	96
(31^8)	8764	3108	4718	2422	1078	19040	9772	5012	3444	1764	938	210
(321^6)	5437	2006	3023	1585	736	11530	6058	3179	2229	1167	643	154
(32^21^4)	3365	1297	1924	1036	500	6890	3719	2005	1433	771	439	113
(32^31^2)	2078	840	1216	677	338	4070	2263	1259	914	509	298	83
(32^4)	1280	544	763	443	227	2380	1366	788	578	336	201	61
(3^21^5)	2686	1044	1584	848	424	5690	3074	1655	1198	641	374	98
(3^221^3)	1655	677	998	553	286	3320	1854	1032	762	422	254	72
(3^22^21)	1018	440	624	361	192	1920	1110	642	480	278	171	53
(3^31^2)	807	355	510	294	162	1530	894	519	398	229	146	46
(3^32)	495	232	315	192	108	870	528	321	247	151	97	34
(41^7)	3843	1414	2177	1134	546	9275	4816	2499	1743	903	497	119
(421^5)	2365	913	1389	743	374	5410	2896	1548	1107	590	339	88
(42^21^3)	1453	591	878	486	254	3125	1730	957	699	386	230	65
(42^31)	891	383	550	318	171	1790	1027	591	438	253	155	48
(431^4)	1151	475	726	399	218	2485	1386	769	575	316	196	57
(4321^2)	705	309	452	260	146	1410	818	473	360	207	132	42
(432^2)	431	201	279	170	97	795	480	291	223	136	88	31
(43^21)	340	163	227	138	82	615	378	231	183	111	75	27
(4^21^3)	484	216	332	189	114	990	580	337	264	151	102	34
(4^221)	295	141	203	123	75	550	337	206	163	99	68	25
(4^23)	141	75	99	65	41	230	151	99	81	53	38	16
(51^6)	1508	582	908	482	254	4051	2140	1129	804	423	243	63
(521^4)	922	377	575	316	174	2276	1250	685	502	274	165	47
(52^21^2)	563	245	360	207	118	1271	727	416	311	178	111	35
(52^3)	343	159	223	136	79	706	421	253	191	116	74	26
(531^3)	443	197	298	170	102	971	565	326	253	144	95	31
(5321)	$\underline{270}$ 1305	129	183	111	68	536	326	198	155	94	63	23
(53^2)	−228	$\underline{69}$ 65	90	59	38	221	144	94	76	50	35	15
(541^2)	534	−63	$\underline{135}$ 390	81	54	351	219	135	111	67	49	19
(542)	−326	37	−183	$\underline{53}$ 160	35	191	125	82	67	44	32	14
(5^21)	−191	26	−128	52	$\underline{26}$ 56	112	76	51	44	29	23	11
(61^5)	174	−22	87	−44	−22	$\underline{1546}$ 51	841	456	336	181	111	32
(621^3)	−685	91	−331	182	89	−170	$\underline{480}$ 608	273	207	117	75	24
(62^21)	518	−72	223	−146	−68	102	−402	$\underline{164}$ 326	126	76	50	18
(631^2)	485	−72	260	−133	−73	102	−378	235	$\underline{102}$ 293	61	43	16
(632)	−366	61	−135	92	52	−51	216	−192	−147	$\underline{40}$ 152	28	12
(641)	−270	37	−198	102	57	−51	200	−132	−168	74	$\underline{22}$ 136	10
(65)	156	−26	93	−52	−41	17	−74	63	63	−52	−52	$\underline{6}$ 41
(71^4)	−150	19	−69	38	16	−45	146	−84	−84	39	39	−11
(721^2)	429	−57	185	−121	−43	108	−381	241	230	−117	−114	33
(72^2)	−116	11	−56	54	11	−27	108	−90	−56	40	36	−11
(731)	−273	46	−114	72	27	−54	199	−114	−159	68	82	−22
(74)	94	−11	61	−50	−11	18	−72	47	61	−22	−50	11
(81^3)	136	−19	62	−31	−16	38	−125	77	70	−39	−32	11
(821)	−276	46	−95	62	27	−57	210	−152	−119	86	54	−22
(83)	138	−35	33	−22	−11	19	−76	57	57	−46	−22	11
(91^2)	−104	11	−62	31	16	−30	101	−61	−62	31	32	−11
(92)	102	−11	42	−40	−11	20	−80	69	42	−40	−22	11
$(10, 1)$	86	−11	53	−22	−16	21	−74	43	53	−22	−32	11
(11)	−66	11	−33	22	11	−11	44	−33	−33	22	22	−11

Table 4.11 (*continued*)

$w=11$ (v)	$h_7h_1^4$	$h_7h_2h_1^2$	$h_7h_2^2$	$h_7h_3h_1$	h_7h_4	$h_8h_1^3$	$h_8h_2h_1$	h_8h_3	$h_9h_1^2$	h_9h_2	$h_{10}h_1$	h_{11}
(1^{11})	7920	3960	1980	1320	330	990	495	165	110	55	11	1
(21^9)	5472	2772	1404	948	246	747	378	129	91	46	10	1
(2^21^7)	3702	1908	983	674	183	552	284	100	74	38	9	1
(2^31^5)	2454	1293	681	475	136	399	210	77	59	31	8	1
(2^41^3)	1596	864	468	332	101	282	153	59	46	25	7	1
(2^51)	1020	570	320	230	75	195	110	45	35	20	6	1
(31^8)	3392	1732	884	604	162	529	269	93	73	37	9	1
(321^6)	2218	1159	605	422	120	379	197	71	58	30	8	1
(32^21^4)	1422	765	411	293	89	265	142	54	45	24	7	1
(32^31^2)	896	499	278	202	66	181	101	41	34	19	6	1
(32^4)	556	322	188	138	49	121	71	31	25	15	5	1
(3^21^5)	1254	670	357	256	78	248	131	49	44	23	7	1
(3^221^3)	778	432	239	176	58	167	92	37	33	18	6	1
(3^22^21)	476	276	160	120	43	110	64	28	24	14	5	1
(3^31^2)	402	234	135	104	38	99	57	25	23	13	5	1
(3^32)	240	147	90	70	28	63	39	19	16	10	4	1
(41^7)	1961	1009	519	358	99	358	183	64	57	29	8	1
(421^5)	1231	653	346	245	73	247	130	48	44	23	7	1
(42^21^3)	759	418	230	167	54	166	91	36	33	18	6	1
(42^31)	461	265	153	113	40	109	63	27	24	14	5	1
(431^4)	647	355	194	143	47	152	82	32	32	17	6	1
(4321^2)	387	223	128	97	35	98	56	24	23	13	5	1
(432^2)	229	139	85	65	26	62	38	18	16	10	4	1
(43^21)	187	115	70	56	23	54	33	16	15	9	4	1
(4^21^3)	304	174	99	76	28	86	48	20	22	12	5	1
(4^221)	176	107	65	51	21	53	32	15	15	9	4	1
(4^23)	80	53	35	29	14	27	18	10	9	6	3	1
(51^6)	1045	544	283	198	57	229	118	42	43	22	7	1
(521^4)	627	340	184	133	42	151	81	31	32	17	6	1
(52^21^2)	371	211	120	89	31	97	55	23	23	13	5	1
(52^3)	217	130	79	59	23	61	37	17	16	10	4	1
(531^3)	303	173	98	75	27	86	48	20	22	12	5	1
(5321)	175	106	64	50	20	53	32	15	15	9	4	1
(53^2)	79	52	34	28	13	27	18	10	9	6	3	1
(541^2)	128	78	47	38	16	44	26	12	14	8	4	1
(542)	72	47	31	25	12	26	17	9	9	6	3	1
(5^21)	48	32	21	18	9	20	13	7	8	5	3	1
(61^5)	501	266	141	101	31	136	71	26	31	16	6	1
(621^3)	287	161	90	67	23	85	47	19	22	12	5	1
(62^21)	163	97	58	44	17	52	31	14	15	9	4	1
(631^2)	127	77	46	37	15	44	26	12	14	8	4	1
(632)	71	46	30	24	11	26	17	9	9	6	3	1
(641)	48	32	21	18	9	20	13	7	8	5	3	1
(65)	16	12	9	8	5	8	6	4	4	3	2	1
(71^4)	$\frac{\mathbf{209}}{\mathbf{45}}$	115	63	47	16	73	39	15	21	11	5	1
(721^2)	−108	$\frac{\mathbf{68}}{\mathbf{294}}$	40	31	12	43	25	11	14	8	4	1
(72^2)	27	−88	$\frac{\mathbf{26}}{\mathbf{50}}$	20	9	25	16	8	9	6	3	1
(731)	54	−159	36	$\frac{\mathbf{17}}{\mathbf{127}}$	8	20	13	7	8	5	3	1
(74)	−18	61	−25	−50	$\frac{\mathbf{5}}{\mathbf{39}}$	8	6	4	4	3	2	1
(81^3)	−38	87	−20	−40	11	$\frac{\mathbf{34}}{\mathbf{38}}$	19	8	13	7	4	1
(821)	57	−153	40	78	−22	−57	$\frac{\mathbf{12}}{\mathbf{112}}$	6	8	5	3	1
(83)	−19	57	−11	−46	11	19	−46	$\frac{\mathbf{4}}{\mathbf{35}}$	4	3	2	1
(91^2)	30	−71	20	32	−11	−30	41	−11	$\frac{\mathbf{7}}{\mathbf{30}}$	4	3	1
(92)	−20	60	−29	−22	11	20	−40	11	−20	$\frac{\mathbf{3}}{\mathbf{29}}$	2	1
$(10, 1)$	−21	53	−11	−32	11	21	−32	11	−21	11	$\frac{\mathbf{2}}{\mathbf{21}}$	1
(11)	11	−33	11	22	−11	−11	22	−11	11	−11	−11	$\frac{\mathbf{1}}{\mathbf{11}}$

Table 4.12

$w = 12$ (i)	h_1^{12}	$h_2h_1^{10}$	$h_2^2h_1^8$	$h_2^3h_1^6$	$h_2^4h_1^4$	$h_2^5h_1^2$	h_2^6	$h_3h_1^9$	$h_3h_2h_1^7$	$h_3h_2^2h_1^5$
(1^{12})	479001600 **1**	239500800	119750400	59875200	29937600	14968800	7484400	79833600	39916800	19958400
(21^{10})	−11	121564800 **122**	61689600	31298400	15876000	8051400	4082400	41731200	21168000	10735200
(2^21^8)	45	−504	31772160 **2107**	16359840	8421840	4334400	2230200	21772800	11208960	5769120
(2^31^6)	−84	952	−4039	8550900 **7890**	4469040	2335500	1220400	11340000	5927040	3097620
(2^41^4)	70	−805	3480	−6970	2372256 **6376**	1259640	669060	5896800	3129840	1661760
(2^51^2)	−21	246	−1090	2260	−2178	680070 **812**	367560	3061800	1650600	890700
(2^6)	1	−12	55	−120	126	−56	202410 **7**	1587600	869400	477000
(31^9)	10	−111	458	−861	720	−215	10	15240960 **103**	7922880	4112640
(321^7)	−72	808	−3378	6452	−5500	1680	−80	−750	4184040 **5545**	2206680
(32^21^5)	168	−1911	8124	−15846	13872	−4380	216	1767	−13314	1183000 **32770**
(32^31^3)	−140	1620	−7040	14136	−12872	4288	−224	−1480	11418	−29044
(32^41)	30	−355	1590	−3332	3236	−1197	70	315	−2500	6636
(3^21^6)	28	−315	1317	−2504	2106	−624	28	300	−2225	5330
(3^221^4)	−105	1200	−5115	9958	−8614	2628	−120	−1140	8654	−21363
$(3^22^21^2)$	90	−1050	4596	−9264	8385	−2694	126	984	−7698	19830
(3^22^3)	−10	120	−546	1164	−1149	426	−20	−108	876	−2388
(3^31^3)	20	−230	984	−1911	1626	−474	20	226	−1734	4299
(3^321)	−20	236	−1044	2118	−1902	576	−24	−228	1821	−4788
(3^4)	1	−12	54	−111	99	−27	1	12	−99	270
(41^8)	−9	100	−413	778	−655	200	−10	−93	675	−1584
(421^6)	56	−630	2641	−5066	4362	−1372	72	586	−4326	10364
(42^21^4)	−105	1200	−5130	10088	−8972	2952	−168	−1110	8374	−20633
(42^31^2)	60	−700	3074	−6268	5867	−2086	140	636	−4932	12630
(42^4)	−5	60	−274	592	−609	260	−30	−52	416	−1116
(431^5)	−42	474	−1988	3796	−3224	984	−48	−454	3364	−8037
(4321^3)	120	−1380	5924	−11642	10242	−3264	168	1316	−10024	24810
(432^21)	−60	708	−3144	6468	−6060	2100	−120	−660	5214	−13605
(43^21^2)	−30	348	−1503	2952	−2556	780	−36	−345	2667	−6651
(43^22)	12	−144	651	−1362	1287	−434	20	138	−1122	3024
(4^21^4)	15	−170	716	−1375	1183	−375	21	164	−1214	2890
(4^221^2)	−30	348	−1509	3006	−2709	918	−60	−333	2550	−6338
(4^22^2)	6	−72	327	−696	694	−280	30	66	−528	1400
(4^231)	12	−141	618	−1236	1100	−358	20	141	−1104	2786
(4^3)	−1	12	−54	112	−106	40	−4	−12	96	−248
(51^7)	8	−89	368	−694	584	−176	8	83	−603	1420
(521^5)	−42	474	−1994	3840	−3316	1032	−48	−442	3275	−7901
(52^21^3)	60	−690	2972	−5899	5300	−1740	84	638	−4852	12115
(52^31)	−20	236	−1052	2190	−2110	772	−40	−212	1667	−4371
(531^4)	30	−340	1432	−2745	2334	−699	30	328	−2443	5882
(5321^2)	−60	696	−3018	6000	−5334	1680	−72	−666	5136	−12946
(532^2)	12	−144	654	−1389	1364	−502	20	132	−1065	2878
(53^21)	12	−141	618	−1233	1078	−316	12	141	−1113	2857
(541^3)	−20	228	−966	1866	−1612	504	−24	−222	1654	−3972
(5421)	24	−282	1242	−2522	2322	−788	40	270	−2104	5368
(543)	−6	72	−324	670	−620	204	−8	−72	582	−1540
(5^21^2)	6	−69	295	−575	500	−154	6	68	−512	1246
(5^22)	−3	36	−163	344	−335	122	−4	−34	272	−724
(61^6)	−7	78	−323	610	−514	156	−8	−73	531	−1252
(621^4)	30	−340	1437	−2782	2417	−762	42	318	−2368	5746
(62^21^2)	−30	348	−1515	3048	−2787	942	−60	−321	2469	−6259
(62^3)	4	−48	219	−472	482	−200	20	42	−336	904
(631^3)	−20	228	−966	1862	−1590	480	−24	−222	1666	−4039
(6321)	24	−282	1242	−2516	2284	−732	40	270	−2122	5486
(63^2)	−3	36	−162	334	−303	90	−4	−36	294	−792
(641^2)	12	−138	590	−1150	1003	−322	20	136	−1024	2480
(642)	−6	72	−326	688	−672	256	−24	−68	544	−1440
(651)	−6	70	−304	602	−530	164	−8	−70	538	−1338
(6^2)	1	−12	54	−112	105	−36	3	12	−96	252
(71^5)	6	−67	278	−526	444	−135	6	63	−459	1084
(721^3)	−20	228	−970	1892	−1658	528	−24	−214	1605	−3927
(72^21)	12	−141	624	−1283	1208	−422	20	129	−1009	2621
(731^2)	12	−138	590	−1147	986	−300	12	136	−1033	2532
(732)	−6	72	−326	686	−658	228	−8	−68	550	−1484
(741)	−6	70	−304	602	−532	172	−8	−70	538	−1330
(75)	2	−24	108	−224	210	−71	2	24	−192	504
(81^4)	−5	56	−233	442	−374	114	−6	−53	387	−916
(821^2)	12	−138	593	−1170	1039	−338	20	130	−986	2444
(82^2)	−3	36	−164	352	−356	144	−12	−32	256	−688
(831)	−6	70	−304	600	−520	156	−8	−70	544	−1367
(84)	2	−24	108	−224	211	−76	6	24	−192	500
(91^3)	4	−45	188	−358	304	−92	4	43	−315	748
(921)	−6	70	−306	616	−558	180	−8	−66	511	−1303
(93)	2	−24	108	−223	204	−63	2	24	−195	522
$(10, 1^2)$	−3	34	−143	274	−235	74	−4	−33	243	−577
$(10, 2)$	2	−24	109	−232	230	−88	6	22	−176	472
$(11, 1)$	2	−23	98	−189	160	−47	2	23	−173	417
(12)	−1	12	−54	112	−105	36	−2	−12	96	−252

Table 4.12 (*continued*)

$w = 12$ (ii)	$h_3h_2^3h_1^3$	$h_3h_2^4h_1$	$h_3^2h_1^6$	$h_3^2h_2h_1^4$	$h_3^2h_2^2h_1^2$	$h_3^2h_2^3$	$h_3^3h_1^3$	$h_3^3h_2h_1$	h_3^4	$h_4h_1^8$	$h_4h_2h_1^6$
(1^{12})	9979200	4989600	13305600	6652800	3326400	1663200	2217600	1108800	369600	19958400	9979200
(21^{10})	5443200	2759400	7257600	3679200	1864800	945000	1260000	638400	218400	10886400	5518800
(2^21^8)	2968560	1527120	3951360	2032800	1045520	537600	715680	367920	129360	5906880	3039120
(2^31^6)	1618740	845820	2147040	1121940	586200	306240	406260	212220	76800	3190320	1667250
(2^41^4)	882576	468888	1164240	618480	328656	174696	230400	122496	45696	1716120	911520
(2^51^2)	481140	260190	630000	340500	184240	99810	130500	70740	27240	919800	496800
(2^6)	262260	144540	340200	187200	103260	57120	73800	40860	16260	491400	270000
(31^9)	2131920	1103760	2862720	1481760	766080	395640	530880	273840	97440	4354560	2252880
(321^7)	1162350	611520	1554000	817600	429660	225540	301560	158130	57960	2338560	1230600
(32^21^5)	633660	339120	841680	450540	240960	128760	171120	91380	34560	1251600	670170
(32^31^3)	345405 27013	188256	454860	247920	135114	73626	96960	52833	20652	667800	363960
(32^41)	−6630	104628 1889	245280	136220	75744	42174	54840	30552	12360	355320	197160
(3^21^6)	−4506	948	608320 926	328240	176720	94940	127200	68220	26160	907200	490560
(3^221^4)	18750	−4098	−3606	180516 14613	99116	54332	72108	39492	15672	482160	265570
$(3^22^21^2)$	−18513	4350	3156	−13482	55576 13627	31150	40788	22866	9408	255640	143440
(3^22^3)	2466	−732	−339	1521	−1702	17896 316	23016	13236	5652	135240	77310
(3^31^3)	−3714	762	762	−3141	2852	−292	30372 745	17130	7168	183120	104010
(3^321)	4509	−974	−783	3492	−3693	402	−808	9929 1183	4324	96600	55920
(3^4)	−261	48	45	−216	252	−21	56	−105	2008 15	36120	21600
(41^8)	1326	−288	−269	1008	−858	96	−195	189	−9	1421280 87	753480
(421^6)	−8900	2000	1731	−6646	5823	−678	1302	−1302	63	−548	407415 3523
(42^21^4)	18383	−4350	−3340	13254	−12120	1500	−2613	2745	−135	1038	−6848
(42^31^2)	−11910	3072	1926	−7968	7790	−1066	1534	−1758	87	−600	4093
(42^4)	1136	−340	−153	654	−674	100	−116	136	−6	52	−374
(431^5)	6798	−1473	−1404	5385	−4626	513	−1116	1080	−54	424	−2741
(4321^3)	−21906	5004	4188	−16830	15246	−1804	3582	−3686	192	−1214	8098
(432^21)	13006	−3310	−2118	9053	−9146	1250	−1864	2274	−126	606	−4203
(43^21^2)	5777	−1250	−1188	4888	−4323	474	−1182	1143	−63	309	−2100
(43^22)	−2978	748	477	−2156	2328	−318	486	−678	45	−120	852
(4^21^4)	−2446	546	509	−1911	1605	−181	384	−348	15	−161	1048
(4^221^2)	5638	−1346	−1069	4252	−3756	450	−894	834	−36	324	−2200
(4^22^2)	−1376	380	210	−892	880	−112	176	−192	9	−66	476
(4^231)	−2448	562	500	−2066	1776	−204	516	−432	18	−132	922
(4^3)	224	−56	−44	184	−152	16	−48	32	−1	12	−88
(51^7)	−1198	260	239	−901	786	−90	171	−177	9	−78	488
(521^5)	6875	−1557	−1303	5049	−4569	555	−966	1044	−54	415	−2657
(52^21^3)	−11051	2694	1916	−7721	7381	−998	1473	−1689	87	−599	3952
(52^31)	4316	−1234	−633	2669	−2764	482	−478	574	−24	203	−1395
(531^4)	−5041	1086	1021	−3955	3529	−399	802	−860	48	−307	1979
(5321^2)	11758	−2724	−2138	8799	−8468	1064	−1825	2148	−126	612	−4094
(532^2)	−2946	884	417	−1831	1998	−390	342	−432	18	−123	865
(53^21)	−2577	538	499	−2147	2091	−222	519	−633	45	−123	842
(541^3)	3417	−770	−696	2631	−2299	274	−511	514	−24	220	−1432
(5421)	−4988	1252	874	−3578	3424	−472	716	−786	36	−264	1810
(543)	1460	−360	−262	1150	−1140	156	−280	313	−15	66	−470
(5^21^2)	−1095	250	217	−830	760	−97	155	−174	9	−68	444
(5^22)	730	−220	−109	460	−485	99	−80	93	−3	34	−238
(61^6)	1053	−225	−211	801	−696	75	−156	162	−9	69	−432
(621^4)	−5012	1116	945	−3708	3377	−380	728	−812	48	−300	1931
(62^21^2)	5802	−1394	−971	4009	−3958	484	−776	984	−63	303	−2024
(62^3)	−928	280	123	−530	566	−92	92	−120	6	−42	298
(631^3)	3463	−730	−706	2774	−2481	258	−585	642	−42	208	−1353
(6321)	−5128	1156	884	−3788	3888	−432	798	−1122	90	−246	1674
(63^2)	762	−156	−132	612	−684	66	−151	255	−30	30	−210
(641^2)	−2130	472	440	−1688	1455	−156	348	−342	18	−136	900
(642)	1408	−368	−222	948	−971	118	−186	244	−15	68	−484
(651)	1190	−259	−234	930	−874	95	−188	229	−15	70	−468
(6^2)	−240	54	42	−180	189	−18	38	−60	6	−12	84
(71^5)	−913	201	183	−696	597	−70	138	−132	6	−60	376
(721^3)	3459	−810	−643	2547	−2303	294	−512	520	−24	203	−1317
(72^21)	−2528	674	391	−1654	1688	−260	316	−384	18	−123	836
(731^2)	−2182	484	447	−1784	1560	−186	399	−378	18	−127	839
(732)	1492	−408	−224	1001	−1102	182	−203	283	−15	62	−434
(741)	1157	−271	−238	941	−795	99	−213	174	−6	70	−476
(75)	−485	130	84	−355	355	−59	75	−83	3	−24	168
(81^4)	773	−166	−155	591	−513	56	−117	118	−6	51	−320
(821^2)	−2178	504	397	−1602	1475	−176	328	−348	18	−124	815
(82^2)	704	−216	−95	410	−431	78	−74	84	−3	32	−226
(831)	1195	−247	−242	1010	−917	91	−243	253	−15	64	−433
(84)	−464	116	86	−364	332	−40	88	−80	3	−24	172
(91^3)	−637	138	127	−483	429	−48	93	−102	6	−42	264
(921)	1208	−283	−205	859	−858	105	−172	223	−15	63	−421
(93)	−501	114	87	−396	423	−51	97	−138	12	−21	147
$(10, 1^2)$	490	−111	−100	375	−321	39	−72	69	−3	33	−210
$(10, 2)$	−480	140	67	−290	315	−54	50	−68	3	−22	154
$(11, 1)$	−350	71	75	−290	246	−24	62	−59	3	−23	150
(12)	240	−60	−42	180	−180	24	−40	48	−3	12	−84

Table 4.12 (*continued*)

$w = 12$ (iii)	$h_4h_2^2h_1^4$	$h_4h_2^3h_1^2$	$h_4h_2^4$	$h_4h_3h_1^5$	$h_4h_3h_2h_1^3$	$h_4h_3h_2^2h_1$	$h_4h_3^2h_1^2$	$h_4h_3^2h_2$	$h_4^2h_1^4$	$h_4^2h_2h_1^2$	$h_4^2h_2^2$	$h_4^2h_3h_1$
(1^{12})	4989600	2494800	1247400	3326400	1663200	831600	554400	277200	831600	415800	207900	138600
(21^{10})	2797200	1417500	718200	1890000	957600	485100	327600	165900	491400	248850	126000	85050
(2^21^8)	1563240	803880	413280	1070160	550200	282800	193480	99400	289380	148680	76370	52220
(2^31^6)	871200	455175	237780	603810	315450	164775	114180	59625	169740	88650	46290	32070
(2^41^4)	484308	257400	136842	339480	180468	95964	67308	35808	99144	52740	28062	19692
(2^51^2)	268620	145410	78810	190200	103020	55870	39620	21530	57660	31305	17020	12085
(2^6)	148680	82080	45450	106200	58680	32520	23280	12960	33390	18540	10335	7410
(31^9)	1164240	601020	309960	806400	415800	214200	147840	76020	224280	115290	59220	40740
(321^7)	646800	339570	178080	454020	238105	124740	87290	45640	131460	68740	35910	25060
(32^21^5)	358540	191655	102360	254660	136040	72610	51480	27435	76660	40870	21770	15410
(32^31^3)	198348	108084	58890	142320	77550	42252	30312	16512	44484	24234	13200	9468
(32^41)	109524	60918	33930	79260	44108	24582	17812	9950	25692	14333	8010	5810
(3^21^6)	264600	142380	76440	190840	102490	54920	39400	21030	59380	31720	16910	12080
(3^221^4)	146040	80175	43940	106350	58336	31941	23210	12671	34340	18776	10246	7432
$(3^22^21^2)$	80468	45128	25298	59060	33128	18574	13640	7644	19764	11084	6212	4564
(3^22^3)	44268	25395	14598	32690	18770	10801	7994	4617	11324	6527	3772	2797
(3^31^3)	58980	33375	18840	43920	24858	14028	10458	5877	15180	8568	4818	3594
(3^321)	32388	18762	10866	24240	14061	8154	6126	3554	8664	5034	2922	2202
(3^4)	12948	7776	4674	9840	5928	3576	2740	1660	3744	2268	1374	1060
(41^8)	398580	210420	110880	284760	149940	78820	55860	29260	86310	45150	23590	16590
(421^6)	219810	118350	63600	159195	85510	45850	32970	17615	50310	26860	14320	10230
(42^21^4)	$\underline{121027}$ 13791	66537	36528	88640	48647	26663	19427	10619	29140	15922	8690	6302
(42^31^2)	−8664	$\underline{37398}$ 5908	21018	49170	27609	15503	11421	6410	16782	9408	5276	3876
(42^4)	864	−700	$\underline{12126}$ 140	27180	15632	9014	6696	3874	9615	5543	3209	2379
(431^5)	5308	−3099	264	$\underline{66130}$ 2246	36560	20145	14890	8155	22570	12370	6760	4960
(4321^3)	−16376	10092	−920	−6622	$\underline{20713}$ 20777	11707	8757	4929	12930	7288	4098	3054
(432^21)	9022	−6134	640	3328	−11202	$\underline{6804}$ 6988	5133	2983	7372	4282	2488	1875
(43^21^2)	4299	−2582	212	1840	−6019	3120	$\underline{3941}$ 2076	2299	5630	3302	1928	1482
(43^22)	−1874	1286	−112	−706	2510	−1668	−780	$\underline{1395}$ 498	3182	1930	1172	906
(4^21^4)	−2040	1215	−116	−872	2521	−1258	−673	244	$\underline{8512}$ 372	4816	2714	2052
(4^221^2)	4550	−2930	328	1818	−5738	3108	1644	−624	−764	$\underline{2819}$ 1808	1646	1263
(4^22^2)	−1088	848	−156	−368	1280	−848	−336	156	164	−464	$\underline{1002}$ 207	774
(4^231)	−1952	1228	−120	−828	2800	−1464	−1024	337	328	−868	202	$\underline{614}$ 595
(4^3)	200	−144	24	80	−288	160	112	−28	−36	112	−42	−84
(51^7)	−917	524	−42	−374	1062	−534	−261	108	140	−272	52	104
(521^5)	5135	−3039	252	2052	−6020	3164	1500	−658	−776	1566	−312	−612
(52^21^3)	−7953	4970	−436	−3035	9355	−5276	−2346	1128	1155	−2480	528	992
(52^31)	2980	−2042	200	1036	−3406	2160	816	−456	−408	952	−224	−376
(531^4)	−3806	2195	−168	−1627	4740	−2413	−1259	532	630	−1246	232	518
(5321^2)	8290	−5088	400	3343	−10509	5854	2913	−1386	−1253	2734	−544	−1232
(532^2)	−1890	1306	−104	−671	2302	−1530	−594	360	254	−607	122	275
(53^21)	−1735	1042	−72	−742	2464	−1350	−836	388	258	−608	112	360
(541^3)	2771	−1630	136	1205	−3426	1736	860	−352	−524	1008	−192	−384
(5421)	−3780	2448	−224	−1494	4756	−2731	−1272	599	632	−1437	312	618
(543)	1020	−660	48	422	−1480	873	527	−253	−160	413	−76	−274
(5^21^2)	−855	495	−35	−381	1065	−546	−249	112	170	−302	50	100
(5^22)	510	−340	22	194	−630	394	147	−81	−85	187	−28	−71
(61^6)	817	−474	42	327	−942	479	236	−98	−120	242	−52	−94
(621^4)	−3778	2296	−220	−1473	4401	−2370	−1109	502	549	−1154	272	448
(62^21^2)	4172	−2750	308	1519	−4806	2870	1182	−630	−575	1298	−368	−484
(62^3)	−676	532	−96	−208	712	−496	−156	96	89	−236	102	76
(631^3)	2639	−1558	136	1115	−3317	1722	903	−390	−427	882	−192	−368
(6321)	−3492	2280	−224	−1342	4367	−2667	−1165	677	491	−1117	308	446
(63^2)	450	−294	24	180	−627	405	198	−135	−57	135	−36	−63
(641^2)	−1782	1087	−114	−766	2241	−1146	−593	233	338	−695	164	273
(642)	1080	−810	132	384	−1292	856	314	−184	−171	440	−180	−152
(651)	930	−561	48	408	−1193	635	298	−143	−183	349	−72	−118
(6^2)	−180	126	−18	−72	228	−144	−54	36	33	−72	27	18
(71^5)	−712	408	−32	−285	834	−419	−221	86	102	−212	40	94
(721^3)	2601	−1570	128	1008	−3099	1666	839	−358	−365	802	−160	−368
(72^21)	−1764	1180	−104	−620	2049	−1289	−525	295	229	−536	122	235
(731^2)	−1663	970	−70	−708	2203	−1122	−680	260	261	−584	104	312
(732)	942	−640	44	343	−1191	809	326	−223	−119	281	−56	−134
(741)	971	−589	48	419	−1313	663	415	−139	−181	415	−76	−226
(75)	−355	225	−12	−149	485	−284	−142	71	65	−142	18	71
(81^4)	607	−352	32	243	−708	356	186	−72	−88	184	−40	−80
(821^2)	−1640	1023	−106	−623	1959	−1065	−543	224	226	−528	132	249
(82^2)	508	−386	60	160	−556	376	138	−72	−65	176	−66	−76
(831)	885	−533	48	373	−1227	640	418	−158	−131	328	−76	−202
(84)	−380	264	−36	−152	528	−304	−184	68	66	−184	58	116
(91^3)	−499	288	−22	−203	580	−300	−142	64	76	−148	28	56
(021)	864	−549	44	320	−1031	612	268	−146	−115	262	−56	−110
(93)	−315	201	−12	−126	447	−279	−153	90	39	−99	18	63
$(10, 1^2)$	400	−234	22	167	−472	237	115	−46	−67	130	−28	−47
$(10, 2)$	−340	250	−32	−112	380	−264	−82	56	45	−112	38	36
$(11, 1)$	−290	164	−12	−127	372	−177	−105	36	52	−105	18	47
(12)	180	−120	12	72	−240	144	72	−36	−30	72	−18	−36

Table 4.12 (*continued*)

$w = 12$ (iv)	h_4^3	$h_5h_1^7$	$h_5h_2h_1^5$	$h_5h_2^2h_1^3$	$h_5h_2^3h_1$	$h_5h_3h_1^4$	$h_5h_3h_2h_1^2$	$h_5h_3h_2^2$	$h_5h_3^2h_1$	$h_5h_4h_1^3$	$h_5h_4h_2h_1$	$h_5h_4h_3$	$h_5^2h_1^2$
(1^{12})	34650	3991680	1995840	997920	498960	665280	332640	166320	110880	166320	83160	27720	33264
(21^{10})	22050	2298240	1164240	589680	298620	398160	201600	102060	68880	103320	52290	17850	21672
(2^21^8)	14070	1310400	673680	346248	177912	236880	121688	62496	42728	63924	32816	11508	14084
(2^31^6)	9000	740880	387090	202212	105615	140130	73182	38211	26472	39366	20550	7428	9120
(2^41^4)	5769	415800	221040	117540	62520	82440	43860	23340	16380	24120	12840	4800	5880
(2^51^2)	3705	231840	125520	68040	36930	48240	26200	14250	10120	14700	8005	3105	3772
(2^6)	2385	128520	70920	39240	21780	28080	15600	8700	6240	8910	4980	2010	2406
(31^9)	11130	1028160	529200	272160	139860	187488	96264	49392	33936	51408	26334	9240	11592
(321^7)	7140	575400	301560	157878	82572	110404	57687	30114	21000	31633	16485	5971	7518
(32^21^5)	4590	320040	170950	91236	48651	64672	34464	18350	12984	19350	10291	3863	4848
(32^31^3)	2955	177030	96450	52545	28623	37686	20529	11181	8019	11766	6408	2502	3108
(32^41)	1905	97440	54180	30168	16824	21848	12192	6816	4944	7112	3981	1622	1980
(3^21^6)	3660	243600	131040	70308	37632	50536	26990	14384	10276	15550	8258	3112	4016
(3^221^4)	2364	133840	73530	40320	22067	29334	16036	8749	6346	9424	5132	2018	2568
$(3^22^21^2)$	1527	73220	41100	23062	12934	16940	9502	5326	3914	5676	3182	1310	1632
(3^22^3)	987	39900	22890	13158	7581	9734	5614	3247	2408	3398	1969	851	1030
(3^31^3)	1230	54600	30990	17544	9903	13092	7386	4152	3096	4524	2541	1059	1344
(3^321)	795	29610	17190	9981	5793	7494	4356	2529	1907	2697	1569	689	846
(3^4)	415	11760	7080	4272	2580	3264	1980	1200	928	1260	768	364	432
(41^8)	4830	393120	205800	107604	56196	75768	39508	20580	14420	22134	11494	4158	5446
(421^6)	3120	215040	115095	61506	32820	44059	23470	12486	8898	13534	7177	2696	3526
(42^21^4)	2019	117180	64150	35079	19161	25492	13907	7579	5487	8214	4466	1750	2264
(42^31^2)	1308	63630	35640	19965	11187	14676	8219	4605	3379	4950	2771	1137	1442
(42^4)	849	34440	19740	11340	6534	8408	4844	2802	2076	2963	1715	739	911
(431^5)	1620	86940	48185	26598	14628	19720	10806	5903	4332	6603	3586	1414	1888
(4321^3)	1053	47005	26675	15099	8525	11311	6374	3583	2670	3958	2218	920	1196
(432^21)	684	25340	14725	8556	4972	6456	3750	2179	1642	2357	1369	599	752
(43^21^2)	555	18550	10940	6435	3773	4942	2897	1691	1299	1868	1090	486	618
(43^22)	360	9940	6005	3633	2201	2797	1697	1031	797	1100	670	317	384
(4^21^4)	729	29190	16870	9684	5526	7474	4238	2392	1814	2780	1554	648	892
(4^221^2)	477	15610	9250	5460	3212	4238	2484	1452	1116	1642	955	423	556
(4^22^2)	312	8330	5060	3074	1870	2392	1452	884	684	964	586	276	344
(4^231)	255	6020	3720	2294	1411	1814	1116	684	542	756	464	225	280
(4^3)	**120** **20**	1890	1230	801	522	648	423	276	225	297	195	105	123
(51^7)	−8	**130872** **74**	70056	37422	19950	27076	14399	7644	5488	8533	4508	1694	2324
(521^5)	48	−393	**38491** **2149**	21102	11544	15541	8483	4621	3380	5172	2805	1103	1498
(52^21^3)	−80	567	−3218	**11886** **5090**	6687	8874	4986	2798	2080	3108	1739	719	956
(52^31)	32	−191	1142	−1974	**3879** **920**	5042	2923	1698	1277	1853	1075	469	604
(531^4)	−40	289	−1598	2375	−801	**6794** **1290**	3845	2166	1640	2493	1395	583	800
(5321^2)	96	−580	3351	−5329	1970	−2676	**2251** **6162**	1314	1009	1476	859	381	502
(532^2)	−20	119	−739	1340	−666	545	−1444	**800** **566**	618	868	528	249	312
(53^21)	−28	115	−678	1085	−358	590	−1447	278	**490** **453**	683	419	203	256
(541^3)	32	−204	1147	−1727	615	−958	1919	−409	−383	**1035** **826**	602	270	378
(5421)	−56	240	−1426	2360	−971	1150	−2664	695	572	−960	**368** **1422**	177	232
(543)	24	−58	358	−620	255	−310	814	−215	−253	227	−400	**95** **202**	114
(5^21^2)	−6	68	−391	600	−215	340	−694	159	133	−306	336	−71	**160** **136**
(5^22)	4	−34	214	−390	198	−170	434	−176	−71	153	−247	59	−68
(61^6)	8	−66	346	−492	161	−249	495	−94	−100	169	−200	48	−53
(621^4)	−40	286	−1551	2295	−786	1135	−2361	470	490	−788	980	−240	255
(62^21^2)	48	−289	1637	−2572	954	−1187	2658	−578	−560	839	−1141	281	−280
(62^3)	−12	38	−228	396	−184	150	−360	100	64	−120	192	−40	35
(631^3)	32	−196	1079	−1585	513	−875	1797	−327	−409	635	−762	199	−216
(6321)	−40	238	−1382	2191	−755	1106	−2573	505	632	−760	1046	−292	277
(63^2)	4	−30	180	−291	87	−159	405	−63	−135	93	−135	51	−36
(641^2)	−28	124	−698	1043	−358	592	−1171	214	248	−516	603	−139	182
(642)	24	−60	360	−598	236	−286	656	−128	−140	244	−368	80	−77
(651)	8	−70	408	−629	211	−366	763	−142	−169	331	−373	83	−148
(6^2)	−2	12	−72	114	−36	66	−144	18	36	−60	72	−12	27
(71^5)	−8	58	−304	438	−151	213	−435	94	88	−139	176	−48	41
(721^3)	32	−195	1067	−1623	605	−758	1646	−397	−340	508	−694	199	−155
(72^21)	−20	119	−692	1153	−501	476	−1145	351	224	−335	514	−150	110
(731^2)	−28	119	−665	1011	−360	534	−1162	262	275	−369	504	−166	116
(732)	8	−62	377	−657	295	−286	741	−251	−161	195	−316	110	−76
(741)	24	−62	359	−568	223	−302	653	−158	−150	259	−363	115	−82
(75)	−4	24	−149	260	−118	130	−319	106	71	−118	177	−59	53
(81^4)	8	−50	262	−375	123	−185	379	−73	−81	118	−148	41	−34
(821^2)	−28	120	−663	1017	−365	471	−1057	227	240	−300	429	−131	86
(82^2)	12	−30	180	−318	156	−114	288	−100	−52	84	−144	40	−23
(831)	24	−60	341	−524	166	−281	657	−115	−193	170	−257	107	−47
(84)	−20	20	−120	200	−80	100	−240	52	68	−80	136	−56	18
(91^3)	−4	42	−222	319	−107	161	−323	65	65	−110	132	−33	34
(921)	8	−63	358	−569	212	−257	605	−139	−137	167	−251	75	−52
(93)	−4	21	−126	210	−75	105	−279	63	90	−57	99	−51	18
$(10, 1^2)$	4	−33	177	−256	89	−134	260	−56	−47	101	−114	24	−34
$(10, 2)$	−4	22	−132	230	−108	90	−224	76	36	−68	112	−24	23
$(11, 1)$	−4	23	−127	186	−59	104	−210	36	47	−81	94	−24	29
(12)	4	−12	72	−120	48	−60	144	−36	−36	48	−72	24	−18

Table 4.12 (*continued*)

$w=12$ (v)	$h_5^2h_2$	$h_6h_1^6$	$h_6h_2h_1^4$	$h_6h_2^2h_1^2$	$h_6h_2^3$	$h_6h_3h_1^3$	$h_6h_3h_2h_1$	$h_6h_3^2$	$h_6h_4h_1^2$	$h_6h_4h_2$	$h_6h_5h_1$	h_6^2	$h_7h_1^5$	$h_7h_2h_1^3$
(1^{12})	16632	665280	332640	166320	83160	110880	55440	18480	27720	13860	5544	924	95040	47520
(21^{10})	10962	408240	206640	104580	52920	70560	35700	12180	18270	9240	3822	672	62640	31680
(2^21^8)	7224	246960	126840	65128	33432	44520	22848	8008	11984	6146	2632	490	40560	20808
(2^31^6)	4758	147510	77040	40227	21000	27870	14547	5257	7821	4080	1809	358	25830	13482
(2^41^4)	3132	87120	46356	24672	13134	17316	9220	3448	5076	2704	1240	262	16200	8628
(2^51^2)	2061	50940	27660	15040	8190	10680	5820	2260	3275	1790	847	192	10020	5460
(2^6)	1356	29520	16380	9120	5100	6540	3660	1480	2100	1185	576	141	6120	3420
(31^9)	5922	206640	105840	54180	27720	37128	18984	6636	10038	5124	2226	420	36000	18360
(321^7)	3906	121800	63560	33138	17262	23079	12012	4340	6538	3395	1533	308	22620	11753
(32^21^5)	2572	71070	37860	20151	10716	14252	7574	2839	4235	2246	1052	226	14000	7436
(32^31^3)	1692	41100	22386	12192	6639	8741	4760	1858	2726	1484	719	166	8550	4656
(32^41)	1113	23580	13148	7344	4110	5324	2982	1216	1743	980	489	122	5160	2888
(3^21^6)	2120	57200	30560	16288	8662	11646	6180	2326	3530	1863	896	196	11960	6338
(3^221^4)	1392	32750	17928	9793	5338	7112	3871	1521	2268	1229	612	144	7210	3928
$(3^22^21^2)$	914	18620	10454	5866	3289	4312	2418	996	1446	810	416	106	4300	2414
(3^22^3)	601	10520	6062	3502	2029	2596	1506	652	915	534	281	78	2540	1472
(3^31^3)	750	14550	8244	4653	2616	3480	1953	813	1200	669	354	92	3540	1998
(3^321)	492	8160	4752	2766	1608	2085	1214	534	756	440	239	68	2070	1209
(3^4)	264	3480	2112	1284	780	988	604	288	388	238	136	44	960	588
(41^8)	2814	93240	48300	25004	12936	17388	8988	3220	4914	2534	1162	238	19080	9804
(421^6)	1860	53265	28250	14966	7920	10614	5610	2093	3174	1672	801	176	11565	6086
(42^21^4)	1226	30210	16431	8929	4848	6443	3495	1363	2037	1102	549	130	6920	3743
(42^31^2)	807	17025	9507	5312	2970	3887	2173	889	1298	726	374	96	4095	2283
(42^4)	531	9540	5474	3152	1824	2330	1348	580	821	479	253	71	2400	1382
(431^5)	1016	23525	12900	7048	3838	5180	2811	1107	1693	911	471	114	5720	3098
(4321^3)	666	13160	7424	4177	2344	3116	1745	723	1076	599	320	84	3345	1874
(432^21)	437	7325	4253	2470	1435	1861	1081	473	678	394	216	62	1940	1126
(43^21^2)	358	5580	3286	1927	1125	1484	865	385	560	324	184	54	1550	910
(43^22)	235	3075	1866	1134	690	876	534	253	348	213	123	40	885	540
(4^21^4)	492	8930	5092	2886	1627	2206	1236	518	800	442	250	68	2510	1406
(4^221^2)	321	4910	2890	1695	991	1310	763	339	502	290	168	50	1430	834
(4^22^2)	210	2690	1634	994	607	772	470	222	312	191	112	37	810	492
(4^231)	171	2010	1244	767	471	610	374	181	256	156	95	32	630	390
(4^3)	81	690	453	297	195	243	159	85	114	75	48	19	240	159
(51^7)	1218	37632	19796	10402	5460	7371	3864	1428	2198	1148	560	126	9276	4817
(521^5)	806	20841	11286	6102	3294	4421	2383	925	1407	755	386	94	5411	2897
(52^21^3)	532	11490	6411	3575	1992	2636	1468	601	893	496	264	70	3126	1731
(52^31)	351	6309	3629	2092	1209	1562	903	391	562	326	179	52	1791	1028
(531^4)	442	8645	4896	2758	1546	2089	1166	485	741	409	228	62	2486	1387
(5321^2)	290	4724	2762	1611	937	1234	717	317	464	268	154	46	1411	819
(532^2)	191	2573	1553	940	571	724	440	207	288	176	103	34	796	481
(53^21)	156	1908	1174	721	441	569	349	169	236	144	88	30	616	379
(541^3)	216	3014	1802	1068	628	850	496	224	344	197	122	38	991	581
(5421)	141	1628	1007	621	382	496	304	147	212	129	81	28	551	338
(543)	75	642	420	275	180	224	147	79	105	69	45	18	231	152
(5^21^2)	96	926	590	372	232	310	192	96	144	87	60	22	352	220
(5^22)	63 77	494	326	215	142	178	117	63	87	57	39	16	192	126
(61^6)	24	13327 64	7186	3865	2074	2820	1509	583	909	483	255	64	4051	2140
(621^4)	-120	-276	4011 1246	2233	1240	1666	923	378	576	317	175	48	2276	1250
(62^21^2)	142	279	-1338	1290 1584	745	977	564	246	361	208	119	36	1271	727
(62^3)	-22	-38	198	-276	451 84	569	344	160	224	137	80	27	706	421
(631^3)	93	186	-858	909	-120	766 675	444	198	299	171	103	32	971	565
(6321)	-127	-228	1137	-1381	200	-875	271 1460	130	184	112	69	24	536	326
(63^2)	12	30	-159	207	-24	138	-279	70 84	91	60	39	16	221	144
(641^2)	-76	-114	543	-593	98	-451	545	-72	136 383	82	55	20	351	219
(642)	34	60	-318	430	-108	244	-408	60	-210	54 220	36	15	191	125
(651)	59	60	-301	350	-48	266	-375	60	-231	120	27 196	12	112	76
(6^2)	-6	-12	66	-90	18	-60	108	-24	54	-48	-48	7 21	32	24
(71^5)	-24	-57	240	-234	30	-156	180	-21	90	-42	-42	6	1546 57	841
(721^3)	103	190	-842	873	-120	561	-695	84	-335	168	164	-24	-190	480 680
(72^21)	-91	-114	543	-633	100	-348	510	-63	215	-128	-118	18	114	-450
(731^2)	-76	-114	516	-518	62	-406	473	-63	257	-114	-129	18	114	-423
(732)	69	57	-285	345	-44	208	-342	51	-117	68	83	-12	-57	242
(741)	59	57	-267	274	-40	224	-233	24	-189	80	94	-12	-57	224
(75)	-47	-19	95	-107	12	-83	107	-12	71	-24	-59	6	19	-83
(81^4)	17	50	-212	213	-30	135	-166	21	-76	42	35	-6	-50	162
(821^2)	-51	-120	543	-594	98	-351	475	-63	204	-134	-94	18	120	-423
(82^2)	22	30	-150	198	-56	84	-128	12	-62	60	24	-6	-30	120
(831)	24	60	-281	303	-40	221	-303	51	-126	80	59	-12	-60	221
(84)	-12	-20	100	-120	28	-80	104	-12	68	-56	-24	6	20	-80
(91^3)	-17	-42	180	-181	22	-119	150	-21	68	-34	-35	6	42	-138
(921)	34	63	-295	337	-44	104	-305	51	-104	68	59	-12	-63	232
(93)	-12	-21	105	-126	12	-84	153	-39	36	-24	-24	6	21	-84
$(10, 1^2)$	17	33	-144	145	-22	101	-114	12	-68	34	35	-6	-33	111
$(10, 2)$	-22	-22	110	-142	32	-68	112	-12	46	-44	-24	6	22	-88
$(11, 1)$	-12	-23	104	-105	12	-81	94	-12	58	-24	-35	6	23	-81
(12)	12	12	-60	72	-12	48	-72	12	-36	24	24	-6	-12	48

Table 4.12 (*continued*)

$w-12$ (vi)	$h_7h_2^2h_1$	$h_7h_3h_1^2$	$h_7h_3h_2$	$h_7h_4h_1$	h_7h_5	$h_8h_1^4$	$h_8h_2h_1^2$	$h_8h_2^2$	$h_8h_3h_1$	h_8h_4	$h_9h_1^3$	$h_9h_2h_1$	h_9h_3	$h_{10}h_1^2$	$h_{10}h_2$	$h_{11}h_1$	h_{12}
(1^{12})	23760	15840	7920	3960	792	11880	5940	2970	1980	495	1320	660	220	132	66	12	1
(21^{10})	16020	10800	5460	2790	582	8460	4275	2160	1455	375	1020	515	175	111	56	11	1
(2^21^8)	10672	7288	3736	1956	428	5910	3028	1551	1058	283	774	396	138	92	47	10	1
(2^31^6)	7035	4872	2541	1365	315	4050	2112	1101	762	213	576	300	108	75	39	9	1
(2^41^4)	4596	3228	1720	948	232	2724	1452	774	544	160	420	224	84	60	32	8	1
(2^51^2)	2980	2120	1160	655	171	1800	985	540	385	120	300	165	65	47	26	7	1
(2^6)	1920	1380	780	450	126	1170	660	375	270	90	210	120	50	36	21	6	1
(31^9)	9360	6384	3252	1704	372	5508	2799	1422	966	255	748	379	130	91	46	10	1
(321^7)	6102	4232	2194	1185	274	3734	1932	999	690	191	553	285	101	74	38	9	1
(32^21^5)	3946	2784	1475	821	202	2482	1314	695	489	143	400	211	78	59	31	8	1
(32^31^3)	2535	1817	989	566	149	1620	882	480	344	107	283	154	60	46	25	7	1
(32^41)	1620	1176	662	388	110	1040	585	330	240	80	196	111	46	35	20	6	1
(3^21^6)	3352	2380	1254	708	176	2246	1180	619	436	127	380	198	72	58	30	8	1
(3^221^4)	2135	1546	837	488	130	1446	783	423	305	95	266	143	55	45	24	7	1
$(3^22^21^2)$	1354	996	558	334	96	916	514	288	212	71	182	102	42	34	19	6	1
(3^22^3)	856	636	372	227	71	572	334	196	146	53	122	72	32	25	15	5	1
(3^31^3)	1122	840	468	288	84	798	447	249	186	63	168	93	38	33	18	6	1
(3^321)	705	534	311	195	62	492	288	168	128	47	111	65	29	24	14	5	1
(3^4)	360	280	172	112	40	252	156	96	76	31	64	40	20	16	10	4	1
(41^8)	5036	3468	1780	954	218	3393	1733	885	605	163	529	269	93	73	37	9	1
(421^6)	3200	2250	1181	657	161	2219	1160	606	423	121	379	197	71	58	30	8	1
(42^21^4)	2023	1451	783	451	119	1423	766	412	294	90	265	142	54	45	24	7	1
(42^31^2)	1274	929	519	308	88	897	500	279	203	67	181	101	41	34	19	6	1
(42^4)	800	590	344	209	65	557	323	189	139	50	121	71	31	25	15	5	1
(431^5)	1673	1216	653	386	104	1255	671	358	257	79	248	131	49	44	23	7	1
(4321^3)	1047	777	432	264	77	779	433	240	177	59	167	92	37	33	18	6	1
(432^21)	654	492	286	179	57	477	277	161	121	44	110	64	28	24	14	5	1
(43^21^2)	531	410	237	154	50	403	235	136	105	39	99	57	25	23	13	5	1
(43^22)	330	256	157	103	37	241	148	91	71	29	63	39	19	16	10	4	1
(4^21^4)	784	590	326	206	62	648	356	195	144	48	152	82	32	32	17	6	1
(4^221^2)	485	372	215	140	46	388	224	129	98	36	98	56	24	23	13	5	1
(4^22^2)	300	232	142	94	34	230	140	86	66	27	62	38	18	16	10	4	1
(4^231)	240	192	117	81	30	188	116	71	57	24	54	33	16	15	9	4	1
(4^3)	105	87	57	42	18	81	54	36	30	15	27	18	10	9	6	3	1
(51^7)	2500	1744	904	498	120	1961	1009	519	358	99	358	183	64	57	29	8	1
(521^5)	1549	1108	591	340	89	1231	653	346	245	73	247	130	48	44	23	7	1
(52^21^3)	958	700	387	231	66	759	418	230	167	54	166	91	36	33	18	6	1
(52^31)	592	439	254	156	49	461	265	153	113	40	109	63	27	24	14	5	1
(531^4)	770	576	317	197	58	647	355	194	143	47	152	82	32	32	17	6	1
(5321^2)	474	361	208	133	43	387	223	128	97	35	98	56	24	23	13	5	1
(532^2)	292	224	137	89	32	229	139	85	65	26	62	38	18	16	10	4	1
(53^21)	232	184	112	76	28	187	115	70	56	23	54	33	16	15	9	4	1
(541^3)	338	265	152	103	35	304	174	99	76	28	86	48	20	22	12	5	1
(5421)	207	164	100	69	26	176	107	65	51	21	53	32	15	15	9	4	1
(543)	100	82	54	39	17	80	53	35	29	14	27	18	10	9	6	3	1
(5^21^2)	136	112	68	50	20	128	78	47	38	16	44	26	12	14	8	4	1
(5^22)	83	68	45	33	15	72	47	31	25	12	26	17	9	9	6	3	1
(61^6)	1129	804	423	243	63	1045	544	283	198	57	229	118	42	43	22	7	1
(621^4)	685	502	274	165	47	627	340	184	133	42	151	81	31	32	17	6	1
(62^21^2)	416	311	178	111	35	371	211	120	89	31	97	55	23	23	13	5	1
(62^3)	253	191	116	74	26	217	130	79	59	23	61	37	17	16	10	4	1
(631^3)	326	253	144	95	31	303	173	98	75	27	86	48	20	22	12	5	1
(6321)	198	155	94	63	23	175	106	64	50	20	53	32	15	15	9	4	1
(63^2)	94	76	50	35	15	79	52	34	28	13	27	18	10	9	6	3	1
(641^2)	135	111	67	49	19	128	78	47	38	16	44	26	12	14	8	4	1
(642)	82	67	44	32	14	72	47	31	25	12	26	17	9	9	6	3	1
(651)	51	44	29	23	11	48	32	21	18	9	20	13	7	8	5	3	1
(6^2)	18	16	12	10	6	16	12	9	8	5	8	6	4	4	3	2	1
(71^5)	456	336	181	111	32	501	266	141	101	31	136	71	26	31	16	6	1
(721^3)	273	207	117	75	24	287	161	90	67	23	85	47	19	22	12	5	1
(72^21)	$\frac{164}{366}$	126	76	50	18	163	97	58	44	17	52	31	14	15	9	4	1
(731^2)	263	$\frac{102}{329}$	61	43	16	127	77	46	37	15	44	26	12	14	8	4	1
(732)	-216	-165	$\frac{40}{172}$	28	12	71	46	30	24	11	26	17	9	9	6	3	1
(741)	-148	-189	83	$\frac{22}{154}$	10	48	32	21	18	9	20	13	7	8	5	3	1
(75)	71	71	-59	-59	$\frac{6}{47}$	16	12	9	8	5	8	6	4	4	3	2	1
(81^4)	-93	-93	43	43	-12	$\frac{209}{50}$	115	63	47	16	73	39	15	21	11	5	1
(821^2)	267	255	-129	-126	36	-120	$\frac{68}{327}$	40	31	12	43	25	11	14	8	4	1
(82^2)	-100	-62	44	40	-12	30	-98	$\frac{26}{56}$	20	9	25	16	8	9	6	3	1
(831)	-126	-177	75	91	-24	60	-177	40	$\frac{17}{142}$	8	20	13	7	8	5	3	1
(84)	52	68	-24	-56	12	-20	68	-28	-56	$\frac{5}{44}$	8	6	4	4	3	2	1
(91^3)	85	77	-43	-35	12	-42	96	-22	-44	12	$\frac{34}{42}$	19	8	13	7	4	1
(921)	-168	-131	95	59	-24	63	-169	44	86	-24	-63	$\frac{12}{124}$	6	8	5	3	1
(93)	63	63	-51	-24	12	-21	63	-12	-51	12	21	-51	$\frac{4}{39}$	4	3	2	1
$(10, 1^2)$	-67	-68	34	35	-12	33	-78	22	35	-12	-33	45	-12	$\frac{7}{33}$	4	3	1
$(10, 2)$	76	46	-44	-24	12	-22	66	-32	-24	12	22	-44	12	-22	$\frac{3}{32}$	2	1
$(11, 1)$	47	58	-24	-35	12	-23	58	-12	-35	12	23	-35	12	-23	12	$\frac{2}{23}$	1
(12)	-36	-36	24	24	-12	12	-36	12	24	-12	-12	24	-12	12	-12	-12	$\frac{1}{12}$

MISCELLANEA

Some procedures for comparing Poisson processes or populations

By Allan Birnbaum,* *Columbia University*

1. Introduction

The problem of comparing two Poisson processes arises in a variety of applications. For example, if a manufactured substance like cloth, paper, or wire is inspected continuously for flaws, the number x of faults observed in any length t may often be assumed to have the probability distribution

$$p(x,\lambda,t) = e^{-\lambda t}(\lambda t)^x/x! \quad (x = 0, 1, 2, \ldots),$$

where λ is the mean number of faults per unit length. Hence the problem of comparing two kinds of material as to flaws is statistically a problem of comparing the parameters λ_1, λ_2 of two Poisson processes.

Alternatively, if a Geiger counter is used to record emissions from a radioactive substance under essentially constant conditions, the number x of emissions recorded during any duration of time t may be assumed to have the probability distribution $p(x,\lambda,t)$, where λ is the mean number of emissions recorded per unit interval of time. The problem of comparing the emission rates of two such substances observed under comparable conditions is statistically again the problem of comparing the parameters λ_1, λ_2 of two Poisson processes.

2. Some experimental designs

A simple experimental procedure which will be convenient for many such problems is the simultaneous observation of the two processes. Assume that two such processes characterized by distributions $p(x_1,\lambda_1,t)$, $p(x_2,\lambda_2,t)$ are observed simultaneously until the first 'event' occurs, i.e. until either $x_1 = 1$ or $x_2 = 1$; in terms of the above examples, an 'event' consists of the observation of a flaw or the recording of a count. (The simultaneous occurrence of two or more 'events', either in the same or in different processes, has probability zero; hence this possibility may be ignored.) Then the probability that the first event occurs in the first process, on the condition that between t and $t+\Delta t$ units of inspection are performed before it occurs, is

$$\lambda_1/(\lambda_1+\lambda_2+\epsilon),$$

where ϵ approaches 0 with Δt; this is seen to be independent of t. Hence if p is the probability that the first event is observed in the first process, $p = \lambda_1/(\lambda_1+\lambda_2) = [(1+(\lambda_2/\lambda_1)]^{-1}$, which depends on λ_1, λ_2 only through their ratio $\gamma = \lambda_2/\lambda_1$. If such simultaneous inspection is continued, and if we set $y_i = 1$ or 0 according as the ith event is observed in the first or second process, the experiment is seen to provide a sequence of independent Bernoulli observations y_i such that

$$\Pr\{y_i = 1\} = p, \qquad \Pr\{y_i = 0\} = 1-p \quad (i = 1, 2, \ldots),$$

where $p = 1/(1+\gamma)$.

For many (but not all) purposes it is appropriate to express comparisons of the two processes in terms of the ratio $\gamma = \lambda_2/\lambda_1$. Statistical questions concerning γ may then be answered by use of the various methods available for dealing with Bernoulli observations.

Problem 1. Suppose that it is required to test whether $\lambda_1 = \lambda_2$ against the alternatives $\lambda_1 \geqslant \lambda_2$, with significance level $\alpha = 0{\cdot}05$, and with power $1-\beta = 0{\cdot}95$ when $\lambda_2 = \frac{1}{2}\lambda_1$. The requirement may be met by any test, based on the observations y_i, of the hypothesis that $p = \frac{1}{2}$ against the alternatives $p > \frac{1}{2}$, with significance level $\alpha = 0{\cdot}05$, and power $1-\beta = 0{\cdot}95$ when $p = \frac{2}{3}$.

Method 1. Using tables of the binomial distribution (or tables of the normal distribution for approximation) we find the smallest value of n such that, for some corresponding critical value r_n, we have

$$\Pr\left\{\sum_{i=1}^{n} y_i > r_n \mid p = \tfrac{1}{2}\right\} \leqslant 0{\cdot}05$$

and

$$\Pr\left\{\sum_{i=1}^{n} y_i > r_n \mid p = \tfrac{2}{3}\right\} \geqslant 0{\cdot}95.$$

---

* This work was begun while the writer was employed by the National Foundation for Infantile Paralysis and was completed with the support of the Office of Naval Research. The writer is indebted to the referee for helpful comments on an earlier draft of this paper and in particular for suggesting the method used in § 3.

The series approximation which was used to obtain Patnaik's table of c and ν for various sample sizes leads to considerable error for small values of ν*. A corrected table of c and ν was given by H. A. David (1951, Table I) but only to the nearest 0·01 in c and 0·1 in ν. This note presents in Table 1 new values of c and ν for the single sample case ($m = 1$) based on recently published moment constants for the distribution of the range in normal samples (Hartley & Pearson, 1951) and an interpolation of the Γ-function at close intervals.

It was noticed that if these corrected values of c and ν are employed, the approximation to the distribution of range by $c\chi/\sqrt{\nu}$ is somewhat better than was indicated by Patnaik (1950). Thus results of Lord (1947), obtained exactly by quadrature, provide factors by which the range (w) in a single sample of n should be multiplied to obtain confidence limits for the mean (μ) of the normal population from which the sample has been drawn. For example, 95 % confidence limits can be obtained from

$$\bar{x} \pm f_{0\cdot05}w,$$

where $\bar{x}$ is the sample mean and $f_{0\cdot05}$ is given in the last column of Table 1 below. If $t_{0\cdot05}$ is the 5 % limit of Student's ratio (two-sided test) for ν degrees of freedom, then $f_{0\cdot05}$ may be approximated by $t_{0\cdot05}/(c\sqrt{n})$. Comparison of the last two columns of the table shows that the agreement is excellent. The differences are negligible for all practical purposes. The agreement is not unexpected in view of Pearson's (1952) findings that the Patnaik χ method provides an excellent approximation to the probability integral of the range for $n = 4, 6, 10, 15$.

Table 1. *Scale factors and equivalent degrees of freedom for χ-approximation to range in normal samples*

No. in sample n	Degrees of freedom ν	Scale factor c	Equivalent two-sided 5 % t	Confidence limits*	
				$\pm t/(c\sqrt{n})$	Lord†
2	1·0000	$1\cdot41421_{4}$	12·7062	6·3531	6·3531
3	$1\cdot9845_{9}$	$1\cdot91154_{7}$	4·3349	1·3093	1·3039
4	$2\cdot9291_{2}$	$2\cdot23886_{6}$	3·2265	0·7206	0·7166
5	$3\cdot8266_{5+}$	$2\cdot48124_{2}$	2·8267	0·5095	0·5066
6	$4\cdot6772_{2}$	$2\cdot67252_{6}$	2·6249	0·4010	0·399
7	$5\cdot4841_{0}$	$2\cdot82980_{6}$	2·5038	0·3344	0·333
8	$6\cdot2511_{7}$	$2\cdot96288_{3}$	2·4233	0·2892	0·288
9	$6\cdot9818_{4}$	$3\cdot07793_{7}$	2·3658	0·2562	0·255
10	$7\cdot6798_{5-}$	$3\cdot17905_{3}$	2·3228	0·2311	0·2301

* Multipliers for sample range to get 95 % confidence limits of the mean.

† From direct quadrature (Lord, 1947).

REFERENCES

David, H. A. (1951). *Biometrika*, **38**, 393.

Hartley, H. O. & Pearson, E. S. (1951). *Biometrika*, **38**, 463.

Lord, E. (1947). *Biometrika*, **34**, 41.

Patnaik, P. B. (1950). *Biometrika*, **37**, 78.

Pearson, E. S. (1952). *Biometrika*, **39**, 130.

---

* Compare the values of c and ν given below with those in Patnaik's Table 1 for $m = 1$.

The third moment of Gini's mean difference

BY A. R. KAMAT, *University College, London*

U. S. Nair (1936) has given an expression for the variance of Gini's mean difference, defined by

$$g = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} |x_i - x_j|. \tag{1}$$

Nair's formula applies to any parental distribution of the x_i, although its actual evaluation may be tedious. Recently Lomnicki (1952) has obtained a simpler expression for the same by a more direct approach. In this note we derive the third moment of g when the parental distribution is normal with the help of the absolute moments of normal distribution obtained by Nabeya (1951, 1952) and Kamat (1953). It may also be observed that this method of derivation is considerably simpler than those adopted by Nair or Lomnicki for deriving the variance of g.

Let x_i ($i = 1, 2, \ldots, n$) denote a sample from a normal population, $N(\mu, \sigma)$. Then if we denote $z_{ij} = x_i - x_j$, z_{ij} is $N(0, \sqrt{2}\sigma)$ and $\rho(z_{ij}, z_{ik}) = \frac{1}{2}$. For the evaluation of the first three moments of g, we require the following expectations which are readily found by using the formulae for absolute moments (see, for example, Kamat, 1953, pp. 26–7):

$$\left.\begin{gathered}
\mathscr{E}(|z_{ij}|) = \frac{2\sigma}{\sqrt{\pi}}, \quad \mathscr{E}(z_{ij}^2) = 2\sigma^2, \quad \mathscr{E}(|z_{ij}||z_{ik}|) = \frac{\sigma^2}{\pi}(2\sqrt{3} + \tfrac{1}{3}\pi) \\
\mathscr{E}(|z_{ij}|^3) = \frac{8\sigma^3}{\sqrt{\pi}}, \quad \mathscr{E}(|z_{ij}|^2|z_{ik}|) = \frac{5\sigma^3}{\sqrt{\pi}}, \\
\mathscr{E}(|z_{ij}||z_{ik}||z_{il}|) = \frac{8\sigma^3}{\pi^{\frac{3}{2}}}\left(\frac{1}{\sqrt{2}} + \tfrac{9}{4}\sin^{-1}\tfrac{1}{3}\right), \\
\mathscr{E}(|z_{ij}||z_{ik}||z_{jk}|) = \frac{3\sigma^3}{\sqrt{\pi}}, \\
\mathscr{E}(|z_{ij}||z_{jk}||z_{kl}|) = \frac{8\sigma^3}{\pi^{\frac{3}{2}}}\left(\frac{1}{\sqrt{2}} - \tfrac{1}{4}\sin^{-1}\tfrac{1}{3} + \sin^{-1}\frac{1}{\sqrt{3}}\right), \\
\mathscr{E}(|z_{ij}||z_{ik}||z_{lp}|) = \frac{8\sigma^3}{\pi^{\frac{3}{2}}}\left(\frac{\sqrt{3}}{2} + \frac{\pi}{12}\right).
\end{gathered}\right\} \tag{2}$$

To find the moments of g the following expectations have to be evaluated:

$$\left.\begin{aligned}
\mathscr{E}(g) &= \frac{2}{n(n-1)}\mathscr{E}(\Sigma|z_{ij}|), \\
\mathscr{E}(g^2) &= \frac{4}{n^2(n-1)^2}\mathscr{E}[\Sigma z_{ij}^2 + 2(\Sigma|z_{ij}||z_{ik}| + \Sigma|z_{ij}||z_{kl}|)], \\
\mathscr{E}(g^3) &= \frac{8}{n^3(n-1)^3}\mathscr{E}[\Sigma|z_{ij}^3| + 3(\Sigma z_{ij}^2|z_{ik}| + \Sigma z_{ij}^2|z_{kl}|) + 6(\Sigma|z_{ij}||z_{ik}||z_{il}| + \Sigma|z_{ij}||z_{ik}||z_{jk}| \\
&\qquad + \Sigma|z_{ij}||z_{jk}||z_{kl}| + \Sigma|z_{ij}||z_{ik}||z_{lp}| + \Sigma|z_{ij}||z_{kl}||z_{pq}|)],
\end{aligned}\right\} \tag{3}$$

where i, j, k, l, p, q are all different and the summation symbols denote summation over all of them with the obvious limitations imposed by the definition (1).

It can be shown that the number of terms in the sums of the various types occurring in (3) are as follows:

Type of sum	*Number of terms*
$\|z_{ij}\|\|z_{ik}\|$	$\frac{1}{2}n(n-1)(n-2)$,
$\|z_{ij}\|\|z_{kl}\|$	$\frac{1}{8}n(n-1)(n-2)(n-3)$,
$z_{ij}^2\|z_{ik}\|$	$n(n-1)(n-2)$,
$z_{ij}^2\|z_{kl}\|$	$\frac{1}{4}n(n-1)(n-2)(n-3)$,
$\|z_{ij}\|\|z_{ik}\|\|z_{il}\|$	$\frac{1}{6}n(n-1)(n-2)(n-3)$,
$\|z_{ij}\|\|z_{ik}\|\|z_{jk}\|$	$\frac{1}{6}n(n-1)(n-2)$,
$\|z_{ij}\|\|z_{jk}\|\|z_{kl}\|$	$\frac{1}{2}n(n-1)(n-2)(n-3)$,
$\|z_{ij}\|\|z_{ik}\|\|z_{lp}\|$	$\frac{1}{4}n(n-1)(n-2)(n-3)(n-4)$,
$\|z_{ij}\|\|z_{kl}\|\|z_{pq}\|$	$\frac{1}{48}n(n-1)(n-2)(n-3)(n-4)(n-5)$.

(4)

Substituting in (3) the values of expectations given in (2) and noting (4), after some simplification, we have the following first three moments of g:

$$\mu_1 = \frac{2\sigma}{\sqrt{\pi}} = (1{\cdot}128379)\,\sigma, \tag{5}$$

$$\begin{aligned}\mu_2 &= \frac{4\sigma^2}{n(n-1)\,\pi}\{(2\sqrt{3}-4+\tfrac{1}{3}\pi)\,n+(6-4\sqrt{3}+\tfrac{1}{3}\pi)\}\\ &= \frac{\sigma^2}{n(n-1)}\{(0{\cdot}651006)\,n+0{\cdot}151508\}\\ &\sim \frac{0{\cdot}651006}{n}\,\sigma^2 \quad \text{for large } n.\end{aligned} \tag{6}$$

$$\begin{aligned}\mu_3 &= \frac{8\sigma^3}{n^2(n-1)^2\,\pi^{\frac{3}{2}}}\{(16\sqrt{2}+40-36\sqrt{3})\,n^2+(144\sqrt{3}-136-80\sqrt{2})\,n+(96\sqrt{2}+120-144\sqrt{3}-2\pi)\}\\ &= \frac{\sigma^3}{n^2(n-1)^2}\{(0{\cdot}393063)\,n^2+(0{\cdot}399735)\,n+0{\cdot}094822\}\\ &\sim \frac{(0{\cdot}393063)\,\sigma^3}{n^2} \quad \text{for large } n.\end{aligned} \tag{7}$$

For $n = 3$ we know that $g = \frac{2}{3}w$, where w is the range. It is easily verified that $\mu_3(g)$ given above checks up with the third moment of w given, for instance, by Hartley & Pearson (1951).

Although it is not possible to say much about the distribution of g on the strength of its first three moments only, the following values of $\sqrt{(\mu_2)}/\mu_1$ and β_1 suggest that it may be close to the χ-distribution. This is confirmed by the χ-approximation to the distribution of the range w for $n = 3$, since in this case $g = \frac{2}{3}w$. The fourth column of the table gives for comparison the values of β_1 for the χ-distribution having the same $\sqrt{(\mu_2)}/\mu_1'$.

n	$\sqrt{(\mu_2)}/\mu_1'$	β_1	β_1 for χ-distribution
5	0·3657	0·1797	0·1672
10	0·2411	0·0708	0·0648
15	0·1926	0·0436	0·0398
20	0·1650	0·0315	0·0286

REFERENCES

Hartley, H. O. & Pearson, E. S. (1951). *Biometrika*, **38**, 463.
Kamat, A. R. (1953). *Biometrika*, **40**, 20.
Lomnicki, Z. A. (1952). *Ann. Math. Statist.* **23**, 635.
Nabeya, S. (1951). *Ann. Inst. Statist. Math.* **3**, 2.
Nabeya, S. (1952). *Ann. Inst. Statist. Math.* **4**, 15.
Nair, U. S. (1936). *Biometrika*, **28**, 428.

A method of systematic sampling based on order properties*

By R. M. SUNDRUM, Ph.D. *Division of Research Techniques, London School of Economics*

1. Introduction

In some theoretical work, particularly in non-parametric inference, the exact probability distributions of certain statistics are so complicated, at least for small samples, that we have to resort to sampling experiments in the laboratory to derive some constants of these distributions. In this paper the method of systematic sampling, which is often employed in surveys and censuses, is utilized to improve upon the method of random samples. In such systematic methods a 'grid' or 'cluster' of points covering the entire population is needed, the randomness entering into the choice of this grid or cluster. It is found that the order properties of samples may be used to determine such a cluster of points.

---

* Part of a thesis submitted for the degree of Ph.D. in the University of London.

One disadvantage lies in the greater difficulty of determining the sampling errors involved. When an idea of the sampling error is not available, the method can be safely used only when there is some external check on the resulting estimates. It has not been possible to evaluate the sampling errors of the method proposed in this paper, but some check is attempted by using some known theoretical results.

2. Systematic sampling from a bivariate population

We consider first the problem of finding the probabilities of the 120 possible rankings of five from a bivariate normal population. In another paper (Sundrum, 1953, submitted to *Biometrika*) these probabilities are used to determine the third and fourth moments of Kendall's rank correlation coefficient in the non-null case. As it appeared extremely difficult to evaluate these probabilities theoretically, it was decided to use estimates based on sampling experiments.

The usual method based on random samples of constructing samples of five from a correlated normal population is to take independent sets, say of five X's and five Y's, from Wold's *Tables of Random Normal Deviates*, and then for each set to find the quantity $Z = \rho X + \sqrt{(1-\rho^2)}\, Y$; then the resulting sets of X and Z may be considered a sample from a bivariate normal population with correlation ρ. The extent to which the sets of X and Z represent samples from a correlated normal population depends on the extent to which the original sets of X and Y represent samples from an independent normal population.

Now, if samples of five bivariate observations of X's and Y's are drawn from an independent normal population, and the X's are arranged in order of magnitude, then it is well known that the 120 possible rankings of the Y's are equally probable. In order to get better estimates for the correlated population, we therefore adopt a method which *ensures* that the samples of correlated observations are constructed from samples of independent observations, when these independent observations are made to conform *exactly* with the condition of equi-probability of the 120 rankings. (In the following a set of X's and Y's will be referred to only by the ranking of the Y's when the X's are arranged in order of magnitude.) The method used is as follows: five X's and five Z's are drawn from Wold's *Tables*. The five Z's are then permuted in the $5! = 120$ possible ways. Each of these sets associated with the set of five X's represents a sample from an independent normal population. If, then, we calculate for each of the 120 bivariate observations the quantities $Y = X + Z'$ (where Z' represents one of the 120 permutations), we get 120 sets of five Y's each, which when associated with the X's may be considered as samples from a bivariate normal population with correlation $\rho = 1/\sqrt{2}$.

To illustrate the method, consider an example with three bivariate observations. From the first six numbers of Wold's *Tables*, write

X	−0·84	1·37	−0·18
Z	0·35	2·82	2·12

The $3! = 6$ samples from an independent population are then

X	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6
−0·84	0·35	0·35	2·12	2·12	2·82	2·82
−0·18	2·12	2·82	0·35	2·82	0·35	2·12
1·37	2·82	2·12	2·82	0·35	2·12	0·35

Then the samples from a correlated normal population are obtained by calculating the values of $Y_i = X + Z_i$, thus:

X	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6
−0·84	−0·49	0·49	1·28	1·28	1·98	1·98
−0·18	1·94	2·64	0·17	2·64	0·17	1·94
1·37	4·19	3·49	4·19	1·72	3·49	1·72

In the same manner, we can also obtain six samples by taking the permutations of the X's. Thus, with a sample of five bivariate observations, we can generate 240 sets of five bivariate observations each, representing samples from a correlated normal population.

In practice, a systematic procedure may be adopted. For the above example, we construct a 3×3 table as follows:

X Z	−0·84	−0·18	1·37
0·35	−0·49	0·17	1·72
2·12	1·28	1·94	3·49
2·82	1·98	2·64	4·19

The body of the table is filled in by adding the values of X and Z corresponding to each cell; e.g. in the first cell on the top left corner we put in $-0{\cdot}84+0{\cdot}35=-0{\cdot}49$. We then get a set of three values, correlated with X, by taking elements from the first, second and third columns, such that no two are in the same row. Similarly, a set of three values, correlated with Y, is obtained by taking elements from the first, second and third rows such that no two are in the same column. This can be done fairly quickly with some practice and may even be mechanized.

This method was used to obtain samples from a bivariate normal population with correlation $\rho = 1/\sqrt{2}$. Fifty bivariate observations of five X's and five Z's were taken from p. 33 of Wold's *Tables*; each of these sets was used to generate 240 observations from the correlated population, giving in all 12,000 bivariate observations. When these 12,000 observations are classified, according to the ranking of the Y's, we obtain the following distribution:

12345	1,044	21345	631	31245	299	41235	123	51234	31
12354	554	21354	274	31254	124	41253	42	51243	25
12435	719	21435	404	31425	161	41325	80	51324	24
12453	282	21453	154	31452	57	41352	29	51342	15
12534	339	21534	132	31524	46	41523	32	51423	16
12543	242	21543	121	31542	32	41532	20	51432	6
13245	631	23145	301	32145	233	42135	92	52134	26
13254	381	23154	133	32154	124	42153	28	52143	16
13425	345	23415	131	32415	110	42315	67	52314	17
13452	129	23451	26	32451	29	42351	10	52341	4
13524	183	23514	53	32514	62	42513	9	52413	11
13542	95	23541	21	32541	24	42531	11	52431	5
14235	416	24135	158	34125	41	43125	63	53124	16
14253	146	24153	38	34152	14	43152	18	53142	5
14325	224	24315	113	34215	52	43215	65	53214	20
14352	95	24351	27	34251	13	43251	15	53241	7
14523	85	24513	16	34512	9	43512	8	53412	7
14532	54	24531	21	34521	6	43521	10	53421	4
15234	179	25134	42	35124	21	45123	19	54123	14
15243	107	25143	27	35142	13	45132	9	54132	5
15324	120	25314	29	35214	29	45213	5	54213	6
15342	58	25341	12	35241	12	45231	5	54231	3
15423	62	25413	9	35412	9	45312	6	54312	4
15432	39	25431	14	35421	4	45321	5	54321	2
				Total	12,000				

From each set of five bivariate observations, we can draw five sets of four bivariate observations, which may also be considered samples from the same correlated population. For example, from the sample represented by (15423), by omitting one observation at a time and re-ranking the remaining observations, we get the samples represented by (4312), (1423), (1423), (1432) and (1432). Thus, we can derive from the above distribution, the distribution of 60,000 observations with four bivariate points each. The distribution of these is as follows:

1234	12,891	2134	6740	3124	2733	4123	847
1243	6,588	2143	3153	3142	944	4132	511
1324	7,571	2314	2788	3214	2129	4213	586
1342	2,825	2341	748	3241	613	4231	292
1423	3,207	2413	1017	3412	395	4312	322
1432	1,979	2431	570	3421	313	4321	238
		Total	60,000				

We can continue the same method to obtain the distribution of rankings of three observations from the above distribution of rankings of four. We then get:

123	106,510
132	48,096
213	47,946
231	13,892
312	14,362
321	9,194
Total	240,000

At this stage we can check the 'quality' of this sample by the correspondence of the above distribution with some theoretical values. I have shown elsewhere that the probabilities of the six rankings of three from a normally correlated population are given by

$$\Pr\{123\} = \frac{3}{8}\left\{\frac{4}{9}+\frac{4}{\pi}\sin^{-1}\rho-\frac{4}{\pi}\sin^{-1}\tfrac{1}{2}\rho+\frac{4}{\pi^2}(\sin^{-1}\rho)^2-\frac{4}{\pi^2}(\sin^{-1}\tfrac{1}{2}\rho)^2\right\}$$

$$\Pr\{132\} = \Pr\{213\} = \frac{3}{4}\left\{\frac{2}{9}-\frac{1}{\pi}\sin^{-1}\rho+\frac{3}{\pi}\sin^{-1}\tfrac{1}{2}\rho-\frac{1}{\pi^2}(\sin^{-1}\rho)^2+\frac{1}{\pi^2}(\sin^{-1}\tfrac{1}{2}\rho)^3\right\}$$

$$\Pr\{231\} = \Pr\{312\} = \frac{3}{4}\left\{\frac{2}{9}+\frac{1}{\pi}\sin^{-1}\rho-\frac{3}{\pi}\sin^{-1}\tfrac{1}{2}\rho-\frac{1}{\pi^2}(\sin^{-1}\rho)^2+\frac{1}{\pi^2}(\sin^{-1}\tfrac{1}{2}\rho)^2\right\}$$

$$\Pr\{321\} = \frac{3}{8}\left\{\frac{4}{9}-\frac{4}{\pi}\sin^{-1}\rho+\frac{4}{\pi}\sin^{-1}\tfrac{1}{2}\rho+\frac{4}{\pi^2}(\sin^{-1}\rho)^2-\frac{4}{\pi^2}(\sin^{-1}\tfrac{1}{2}\rho)^2\right\}.$$

Evaluating these expressions for the case of $\rho = 1/\sqrt{2}$ we can compare with the relative frequencies of these rankings actually achieved in the sample. The results are as follows:

	Theoretical	Sample
123	0·4431	0·4438
132	0·2009	0·2004
213	0·2009	0·1998
231	0·0585	0·0579
312	0·0585	0·0598
321	0·0381	0·0383

The correspondence appears to be satisfactory.

Incidentally, from the distribution of rankings of five, we may derive the small sample distribution of Kendall's rank correlation coefficient from this population. The distribution is as follows:

Distribution of t ($n = 5$) from a normally correlated population $\rho = 1/\sqrt{2}$

t	Frequency
1·0	1044
0·8	2535
0·6	3041
0·4	2452
0·2	1518
0·0	798
−0·2	390
−0·4	149
−0·6	55
−0·8	16
−1·0	2
Total	12,000

The mean and variance of this distribution are 0·4986 (0·50) and 0·10818 (0·10992) respectively, the figures in brackets being the known theoretical values.

In the same manner, we may derive the distribution of the Spearman rank correlation coefficient in samples of five from such a population. The distribution is as follows:

Distribution of r_s $(n = 5)$ from a normally correlated population with $\rho = 1/\sqrt{2}$

r_s	Frequency	r_s	Frequency
1·0	1044	−0·1	235
0·9	2535	−0·2	116
0·8	1059	−0·3	142
0·7	1982	−0·4	55
0·6	1242	−0·5	67
0·5	893	−0·6	55
0·4	562	−0·7	37
0·3	896	−0·8	18
0·2	320	−0·9	16
0·1	505	−1·0	2
0·0	219		
		Total	12,000

The mean and variance of r_s from this distribution are 0·5938 and 0·12365 respectively; the theoretical value of the mean, from Moran's formula, is 0·595.

A final assessment of the value of this method must of course await further investigation on the standard errors involved, and the speed with which each method of sampling can be carried out. Clearly the 12,000 values of, say, r_s obtained by 240 recombinations of each of the fifty independent bivariate samples are 'worth' more than fifty independent values of r_s; but they would probably provide less information than we should obtain by calculating one value of r_s from each of 12,000 independent bivariate samples. Thus if the same accuracy is obtained by drawing N independent samples and calculating one value of the statistic from each, as by drawing n samples each providing m values of the statistic by recombination, the practical problem is to determine which procedure involves the smaller expenditure of effort.

3. Application to the univariate case

The above method of systematic sampling may also be applied to some univariate problems. For example, the sampling distribution of many distribution-free statistics for testing the homogeneity of two samples cannot be determined simply under certain non-null cases. Then we may use a systematic procedure of sampling experiments to obtain such distributions. Let the problem be one of determining the sampling distribution of some statistic based on ranks alone to test the homogeneity of two samples of mX's and nY's, when the X's and Y's are drawn from populations which differ only in location. Then a laboratory sample is obtained by taking mX's and nZ's from identical populations of the specified type, and constructing a sample of nY's by adding, say d, to each of the nZ's. The extent to which the samples of X's and Y's represent the populations with different location depends on the extent to which the X's and Z's represent the identical populations. In order to ensure that the X's and Z's do represent samples from identical populations, they should satisfy the requirement that all the $\binom{m+n}{m}$ ways of partitioning a set of $(m+n)$ observations into two sets of m and n respectively should be equally probable. In order to achieve this exactly, the following suggestion is made. We take a set of $(m+n)$ observations and divide them into two sets of m and n in all the $\binom{m+n}{m}$ possible ways. From each of these sets the n observations are transformed into the second sample of nY's by adding the quantity d to each of them. Such samples may then be used to study the distribution of such statistics as Wilcoxon's two-sample test under alternative hypotheses.

4. Conclusion

In using this method, the practical task is very much simplified because of the systematic nature of the sampling. However, it does not seem practicable to use this method for any large sample, for the task of enumerating all the possible permutations would be prohibitive. The main use of this method seems to be to estimate constants of certain distributions, rather than the distributions themselves.

A note on ordered least-squares estimation

By F. DOWNTON, *University of Liverpool*

The parameters μ, σ of a distribution having the form $f\{(x-\mu)/\sigma\}$ may be estimated by applying the method of least squares to ordered observations, where ordering is according to magnitude. The general theory was given in a paper by Lloyd (1952). It was there shown that the 'ordered' estimate $\hat{\mu}$ of the expectation μ had a variance never exceeding that of the sample mean; and, in the case of symmetric distributions, necessary and sufficient conditions were obtained for the variance of $\hat{\mu}$ to be strictly less than that of the sample mean.

In this note these conditions have been extended to include the unsymmetrical case.

Let $x_1, x_2, \ldots, x_n$ be a sample of n independent observations on a continuous variate X whose distribution has the above form. The possibility of repetitions among the x's then has zero probability. We may therefore, with probability one, write

$$x_{(1)} < x_{(2)} < x_{(3)} < \ldots < x_{(n)} \tag{1}$$

for the ordered observations $x_{(r)}$.

Let

$$z_r = (x_r - \mu)/\sigma, \quad z_{(r)} = (x_{(r)} - \mu)/\sigma,$$

be the reduced observations, unordered and ordered respectively. They have parameter-free distributions, so that their moments may be regarded as known.

Clearly,

$$\mathscr{E}(z_r) = 0 \quad (r = 1, 2, 3, \ldots, n). \tag{2}$$

Let

$$\mathscr{E}(z_{(r)}) = \alpha_r, \quad \operatorname{cov}(z_{(r)}, z_{(s)}) = w_{rs} \quad (r, s = 1, 2, 3, \ldots, n).$$

Let $\boldsymbol{\alpha}$ denote the $(n \times 1)$ vector of the α_r, $\mathbf{w}$ the symmetric, positive-definite $(n \times n)$ matrix of the w_{rs}, $\mathbf{1}$ an $(n \times 1)$ vector of 1's, and $\mathbf{x}$ the $(n \times 1)$ vector of the $x_{(r)}$.

We note for future use that $\mathbf{1}'\boldsymbol{\alpha} = 0$, that $\boldsymbol{\alpha} \neq \mathbf{0}$, and that $\mathbf{1}'\mathbf{w}\mathbf{1} = n$. The first of these results follows from the fact that the z_r and the $z_{(r)}$ are simply permutations of the same set of numbers; hence $\Sigma z_r = \Sigma z_{(r)}$ and

$$\mathbf{1}'\boldsymbol{\alpha} = \Sigma \alpha_r = \Sigma \mathscr{E}(z_{(r)}) = \mathscr{E}(\Sigma z_{(r)}) = \mathscr{E}(\Sigma z_r) = \Sigma \mathscr{E}(z_r) = 0.$$

A similar argument shows that $\mathbf{1}'\mathbf{w}\mathbf{1} = n$.

The second result follows from (1). For with $r < s$, we have $z_{(r)} < z_{(s)}$ and, taking expectations, $\alpha_r < \alpha_s$. Therefore $\alpha_r \neq \alpha_s$ and hence $\boldsymbol{\alpha} \neq \mathbf{0}$.

The 'ordered' estimate $\hat{\mu}$, of μ, and its variance, are given by

$$\hat{\mu} = \boldsymbol{\alpha}'\mathbf{w}^{-1}(\boldsymbol{\alpha}\mathbf{1}' - \mathbf{1}\boldsymbol{\alpha}')\,\mathbf{w}^{-1}\mathbf{x}/\Delta,$$

$$\operatorname{var}(\hat{\mu}) = \boldsymbol{\alpha}'\mathbf{w}^{-1}\boldsymbol{\alpha}\sigma^2/\Delta,$$

where

$$\Delta = (\mathbf{1}'\mathbf{w}^{-1}\mathbf{1})\,(\boldsymbol{\alpha}'\mathbf{w}^{-1}\boldsymbol{\alpha}) - (\mathbf{1}'\mathbf{w}^{-1}\boldsymbol{\alpha})^2.$$

Lloyd showed that when the distribution is symmetric, $\mathbf{1}'\mathbf{w}^{-1}\boldsymbol{\alpha} = 0$, which considerably simplifies these expressions. He also showed that a necessary and sufficient condition for the variance of $\hat{\mu}$ to be strictly smaller than σ^2/n is that $\mathbf{w}\mathbf{1} \neq \mathbf{1}$, or, equivalently, that $\operatorname{var}(\hat{\mu}) = \sigma^2/n$ if and only if $\mathbf{w}\mathbf{1} = \mathbf{1}$. He showed further that when the variance of $\hat{\mu}$ equals that of the sample mean, $\hat{\mu}$ itself coincides with the sample mean.

In the general case, where no symmetry assumptions are made we first note that both $\mathbf{w}$ and $\mathbf{w}^{-1}$ are positive-definite and symmetric and may therefore be expressed in the form

$$\mathbf{w} = \mathbf{t}\mathbf{t}' \quad \text{and} \quad \mathbf{w}^{-1} = (\mathbf{t}^{-1})'\,\mathbf{t}^{-1},$$

where $\mathbf{t}$ is a lower triangular matrix.

Consider any two $(n \times 1)$ vectors $\mathbf{b}$ and $\mathbf{c}$.

Then

$$\mathbf{b}'\mathbf{w}\mathbf{b} = \mathbf{b}'\mathbf{t}\mathbf{t}'\mathbf{b} = \mathbf{h}'\mathbf{h} = \Sigma h_i^2, \quad \text{say},$$

where $\mathbf{h} = \mathbf{t}'\mathbf{b}$.

Also

$$\mathbf{c}'\mathbf{w}^{-1}\mathbf{c} = \mathbf{c}'(\mathbf{t}^{-1})'\,\mathbf{t}^{-1}\mathbf{c} = \mathbf{k}'\mathbf{k} = \Sigma k_i^2, \quad \text{say},$$

where $\mathbf{k} = \mathbf{t}^{-1}\mathbf{c}$.

Now by the Cauchy-Schwarz inequality we have

$$(\Sigma h_i^2)\,(\Sigma k_i^2) \geqslant (\Sigma h_i k_i)^2, \tag{3}$$

and the necessary and sufficient condition for equality is that

$$h_i = \lambda k_i \quad \text{for some constant } \lambda \text{ and for all } i. \tag{4}$$

In matrix form (3) and (4) become

$$(\mathbf{b}'\mathbf{w}\mathbf{b})\,(\mathbf{c}'\mathbf{w}^{-1}\mathbf{c}) \geqslant (\mathbf{b}'\mathbf{c})^2, \tag{5}$$

with

$$\mathbf{b} = \lambda\mathbf{w}^{-1}\mathbf{c} \quad \text{as the condition for equality.} \tag{6}$$

Put $\mathbf{b} = \mathbf{w}^{-1}\mathbf{1} - \mathbf{1}$ and $\mathbf{c} = \boldsymbol{\alpha}$. Then (5) becomes

$$(\mathbf{1}'\mathbf{w}^{-1}\mathbf{1} - 2\mathbf{1}'\mathbf{1} + \mathbf{1}'\mathbf{w}\mathbf{1})\,(\boldsymbol{\alpha}'\mathbf{w}^{-1}\boldsymbol{\alpha}) \geqslant (\mathbf{1}'\mathbf{w}^{-1}\boldsymbol{\alpha} - \mathbf{1}'\boldsymbol{\alpha})^2.$$

Now

$$\mathbf{1}'\mathbf{1} = \mathbf{1}'\mathbf{w}\mathbf{1} = n \quad \text{and} \quad \mathbf{1}'\boldsymbol{\alpha} = 0.$$

Thus

$$(\mathbf{1}'\mathbf{w}^{-1}\mathbf{1} - n)\,(\boldsymbol{\alpha}'\mathbf{w}^{-1}\boldsymbol{\alpha}) \geqslant (\mathbf{1}'\mathbf{w}^{-1}\boldsymbol{\alpha})^2,$$

or

$$\{(\mathbf{1}'\mathbf{w}^{-1}\mathbf{1})\,(\boldsymbol{\alpha}'\mathbf{w}^{-1}\boldsymbol{\alpha}) - (\mathbf{1}'\mathbf{w}^{-1}\boldsymbol{\alpha})^2\}/(\boldsymbol{\alpha}'\mathbf{w}^{-1}\boldsymbol{\alpha}) = \sigma^2/\operatorname{var}(\hat{\mu}) \geqslant n,$$

since $\boldsymbol{\alpha}'\mathbf{w}^{-1}\boldsymbol{\alpha}$ is essentially positive. Thus $\operatorname{var}(\hat{\mu}) \leqslant \sigma^2/n$; and the necessary and sufficient condition for equality is, from (6), that

$$\mathbf{w}\mathbf{1} = \mathbf{1} - \lambda\boldsymbol{\alpha} \tag{7}$$

for some scalar constant λ. When this condition is satisfied not only is the variance of $\hat{\mu}$ equal to that of the sample mean, but $\hat{\mu}$ itself necessarily *is* the sample mean.

For, pre-multiplying (7) by $\boldsymbol{\alpha}'\mathbf{w}^{-1}$, $\mathbf{1}'\mathbf{w}^{-1}$ and $\mathbf{w}^{-1}$ we have

$$\boldsymbol{\alpha}'\mathbf{w}^{-1}\mathbf{1} - \lambda\boldsymbol{\alpha}'\mathbf{w}^{-1}\boldsymbol{\alpha} = \boldsymbol{\alpha}'\mathbf{1} = 0, \tag{8}$$

$$\mathbf{1}'\mathbf{w}^{-1}\mathbf{1} - \lambda\mathbf{1}'\mathbf{w}^{-1}\boldsymbol{\alpha} = \mathbf{1}'\mathbf{1} = n, \tag{9}$$

and

$$\mathbf{w}^{-1}\mathbf{1} - \lambda\mathbf{w}^{-1}\boldsymbol{\alpha} = \mathbf{1}, \tag{10}$$

respectively.

Now

$$\begin{aligned} \Delta &= (\boldsymbol{\alpha}'\mathbf{w}^{-1}\boldsymbol{\alpha})\,(\mathbf{1}'\mathbf{w}^{-1}\mathbf{1}) - (\mathbf{1}'\mathbf{w}^{-1}\boldsymbol{\alpha})^2 \\ &= \boldsymbol{\alpha}'\mathbf{w}^{-1}\boldsymbol{\alpha}(\mathbf{1}'\mathbf{w}^{-1}\mathbf{1} - \lambda\mathbf{1}'\mathbf{w}^{-1}\boldsymbol{\alpha}) \quad \text{from (8)} \\ &= n\boldsymbol{\alpha}'\mathbf{w}^{-1}\boldsymbol{\alpha} \quad \text{from (9)}. \end{aligned} \tag{11}$$

Also

$$\begin{aligned} \hat{\mu} &= (\boldsymbol{\alpha}'\mathbf{w}^{-1}\boldsymbol{\alpha}\mathbf{1}' - \boldsymbol{\alpha}'\mathbf{w}^{-1}\mathbf{1}\boldsymbol{\alpha}')\,\mathbf{w}^{-1}\mathbf{x}/\Delta \\ &= \boldsymbol{\alpha}'\mathbf{w}^{-1}\boldsymbol{\alpha}(\mathbf{1}'\mathbf{w}^{-1} - \lambda\boldsymbol{\alpha}'\mathbf{w}^{-1})\,\mathbf{x}/\Delta \quad \text{from (8)} \\ &= \boldsymbol{\alpha}'\mathbf{w}^{-1}\boldsymbol{\alpha}\mathbf{1}'\mathbf{x}/\Delta \quad \text{from (10)}, \end{aligned}$$

whence $\hat{\mu} = \mathbf{1}'\mathbf{x}/n$, which is the sample mean.

I am grateful to Dr E. H. Lloyd for an extremely stimulating correspondence on this subject.

REFERENCE

Lloyd, E. H. (1952). Least squares estimation of location and scale parameters using order statistics. *Biometrika*, **39**, 88.

A note on the evaluation of the multivariate normal integral

By F. N. DAVID, *University College, London*

1. A problem which arises frequently is concerned with the evaluation of the multivariate normal integral. Thus if $x_1, x_2, \ldots, x_n$ are normal variables, all correlated, we often need

$$P\{x_1 > 0, x_2 > 0, \ldots, x_n > 0\}.$$

Little work of practical use has appeared since the papers of Moran (1948) and Kendall (1941). It is the purpose of this note to challenge others to the solution of the problem and to summarize the position, as the writer sees it, at this present time.

2. Let the event E_i consist in the random variable x_i being positive. Thus

$$P\{x_i > 0\} = P\{E_i\} = \tfrac{1}{2}.$$

For two variables x_1 and x_2 we have the well-known result of Sheppard (1898),

$$P\{x_1 > 0, x_2 > 0\} = P\{E_1 E_2\} = \frac{1}{2}\left[1 - \frac{\cos^{-1}\rho_{12}}{\pi}\right],$$

where ρ_{12} is the correlation between x_1 and x_2. This result is simply achieved by an appeal to geometry as is also the result for three variables. With three variables the probability ellipsoid can be transformed into a sphere with the original co-ordinate planes becoming three planes through the origin. It is easy to show for three (or more) dimensions that the angle between the ith and jth transformed co-ordinate planes is just $\cos^{-1}\rho_{ij}$, where ρ_{ij} is the correlation between the original variables x_i and x_j. The required probability for three (or more) variables is just the solid angle in three or more dimensions when the angles between the planes containing the solid angle are given. An exact expression for this solid angle appears to be known for three dimensions only.

3. A curious recurrence relation for the required probabilities is possible. This relation has its analogue in many branches of mathematics; we may quote, for example, any text-book on multi-dimensional geometry. It was also partially explored in relation to an n-dimensional integral by Schaefli (1857). The theorem usually ascribed to Boole is as follows:

$$P\{E_1 + E_2 + \ldots + E_n\} = \sum_i P\{E_i\} - \sum_{i<j} P\{E_i E_j\} + \sum_{i<j<l} P\{E_i E_j E_l\} - \ldots + (-1)^{n-1} P\{E_1 E_2 \ldots E_n\}.$$

Now for the problem considered

$$P\{E_1 + E_2 + \ldots + E_n\} = 1 - P\{\bar{E}_1 \bar{E}_2 \ldots \bar{E}_n\} = 1 - P\{E_1 E_2 \ldots E_n\},$$

by the symmetry of the multivariate normal distribution. Hence for n *odd* we have

$$P\{E_1 E_2 \ldots E_n\} = \tfrac{1}{2}[1 - \sum_i P\{E_i\} + \sum_{i<j} P\{E_i E_j\} - \ldots \text{etc.}],$$

the last term on the right-hand side being the sum of the joint probabilities of the E's taken $n-1$ at a time. The result for $n = 3$ is immediate:

$$P\{E_1 E_2 E_3\} = \frac{1}{2}\left[1 - \frac{3}{2} + \frac{3}{2} - \frac{\cos^{-1}\rho_{12}}{2\pi} - \frac{\cos^{-1}\rho_{13}}{2\pi} - \frac{\cos^{-1}\rho_{23}}{2\pi}\right]$$

$$= \frac{1}{4\pi}[2\pi - \cos^{-1}\rho_{12} - \cos^{-1}\rho_{13} - \cos^{-1}\rho_{23}].$$

For n even we get an identity.

4. From the recurrence relation we can deduce the results for $n = 5$ given that for $n = 4$, but it does not seem possible to obtain a recurrence relation which will enable us to make the step from 3 to 4. Moran, following Kendall, gave an expansion for $P\{E_1 E_2 E_3 E_4\}$, but experience has shown that this does not converge fast enough for practical purposes except for ρ_{ij} small. For example, for ρ_{ij} in the neighbourhood of 0·5 at least twenty sets of terms are necessary to obtain reasonable accuracy. Some work has been done towards obtaining a series expansion for ρ_{ij} nearly 0·5 in terms of $(\rho_{ij} - 0{\cdot}5)$, but only in certain restricted cases. What is required is a quickly convergent series for ρ_{ij} large and n even. From $n = 4$ we deduce $n = 5$, calculate $n = 6$ from the series, deduce $n = 7$ and so on. The problem is difficult but not unrewarding.

REFERENCES

MORAN, P. A. P. (1948). *Biometrika*, **35**, 203.
KENDALL, M. G. (1941). *Biometrika*, **32**, 196.
SCHAEFLI, L. (1857). *Quart. J. Pure Appl. Math.* **2**, 269.
SHEPPARD, W. F. (1898). *Phil. Trans.* A, **192**, 101.

unequal components there is skewness, and platy- or leptokurtosis (i.e. the excess kurtosis is negative or positive) according as ρ is less or greater than $(2+\sqrt{3})$. If $\rho = (2+\sqrt{3})$ there is no excess kurtosis for any separation. As ρ becomes large, γ_1 and γ_2 both decrease towards zero, and with any finite separation tend to the limit zero as ρ tends to infinity (with infinite separation, γ_2 diverges to infinity). With zero separation, trivially, both γ_1 and γ_2 are zero for all values of ρ. The curves of constant separation are a family of cardioids, and with large values of ρ, near the origin, closely approach the straight lines $\gamma_2 = \gamma_1\delta$. The curves of constant ratio ρ are quartic parabolas, $\gamma_1^4/\gamma_2^3 = \rho(\rho-1)^4/(\rho^2-4\rho+1)^3$ (constant).

4. USE OF THE DIAGRAM

To estimate ρ and δ by the method of moments when we have available a sample of n individuals from the composite population, we obtain in the usual way from the k-statistics estimates g_1 and g_2 of γ_1 and γ_2. These may be used as co-ordinates in the chart to obtain approximate values r and d for ρ and δ, by interpolating between the contours. An estimate, s, of σ may then be obtained from

$$s = (r+1)\sqrt{k_2}/\{(r+1)^2+rd^2\}^{\frac{1}{2}}, \tag{3}$$

while the estimates m_1 and m_2 of μ_1 and μ_2 are given by

$$m_1 = k_1 - ds/(r+1), \quad m_2 = k_1 + rds/(r+1). \tag{4}$$

Here k_1 and k_2 are the first two k-statistics of the composite sample.

5. PRACTICAL USE OF THE METHOD

We take as a first example the data used by Rao (1948) for the same purpose, viz. the heights in centimetres of 454 plants, shown in Table 1.

Table 1. *Heights of* 454 *plants in centimetres*

Central value	8	9	10	11	12	13	14	15	16	17	18	19	20	Total
Frequency	3	9	21	40	59	76	79	69	46	30	13	7	2	454

The k-statistics (calculated by Rao) are:

$$k_1 = -0{\cdot}244 \text{ (about 14 as origin)}, \quad k_3 = 0{\cdot}729,$$
$$k_2 = 4{\cdot}976, \quad k_4 = -5{\cdot}315.$$

Hence

$$g_1 = 0{\cdot}066, \quad g_2 = -0{\cdot}215.$$

Using these values in the chart, we obtain the estimates

$$r = 1{\cdot}4, \quad d = 1{\cdot}4.$$

From equation (3),

$$s = 1{\cdot}836.$$

From equations (4),

$$m_1 = -0{\cdot}244 - 2{\cdot}57/2{\cdot}4 = -1{\cdot}31,$$
$$m_2 = -0{\cdot}244 + 1{\cdot}4 \times 2{\cdot}57/2{\cdot}4 = 1{\cdot}26.$$

The chief source of error in these estimates lies in the difficulty of accurate estimation from the diagram. Assuming a standard error of $\pm 0{\cdot}05$ in the estimation of both r and d, the standard errors of the results due to errors of interpolation may be found from

$$\left.\begin{aligned}
\sigma_{p_1}^2 &= \sigma_r^2/(1+\rho)^4, \quad [p_1 = r/(r+1)],\\
\sigma_s^2 &= \frac{\sigma^6\delta^2}{4\kappa_2^2(\rho+1)^6}\{(\rho-1)^2\delta^2\sigma_r^2 + 4\rho^2(\rho+1)^2\sigma_\alpha^2\},\\
\sigma_{m_1}^2 &= \frac{\sigma^6}{4\kappa_2^2(\rho+1)^6}\{[2(\rho+1)+\delta^2]^2\delta^2\sigma_r^2 + 4(\rho+1)^4\sigma_\alpha^2\},\\
\sigma_{m_2}^2 &= \frac{\sigma^6}{4\kappa_2^2(\rho+1)^6}\{[2(\rho+1)+\rho\delta^2]^2\delta^2\sigma_r^2 + 4\rho^2(\rho+1)^4\sigma_\alpha^2\}.
\end{aligned}\right\} \tag{5}$$

The results of these calculations, together with the theoretical results of Rao by the methods of moments and of maximum likelihood, and the standard errors of the methods due to sampling and assumptions, are given in Table 2.

Table 2

Parameter estimated	Estimate	Estimated value by method of			Standard error of estimation from diagram ($\sigma_r = \sigma_d = 0{\cdot}05$)	Standard error of max. likelihood estimates
		Max. likelihood	Moments	γ_1, γ_2 diagram		
$\rho/(1+\rho)$	p_1	0·567	0·583	0·58	0·009	0·161
μ_1	m_1	−1·345	−1·341	−1·31	0·034	0·049
μ_2	m_2	1·282	1·286	1·26	0·043	0·033
σ	s	1·823	1·816	1·84	0·021	0·049

The reading errors are rather large, but could be reduced if necessary by constructing a large number of contours on as large a scale as was convenient. Also, in some cases the errors would be greater owing to the shape of the diagram, but ρ and δ are usually small ($\rho < 5$, $1 < |\delta| < 3$) and can then be estimated with considerable accuracy.

A second example may be taken from the distributions of length of trypanosome strains analysed in some detail by K. Pearson (1914). The second column of Table 3 shows the observed frequency distribution of the length (in microns) of 1000 individuals of *Trypanosoma gambiense* as recorded by Pearson. The moments and moment ratios of the distribution, using our present notation, are given by Pearson (p. 131)* as

$$k_1 = 22{\cdot}1130, \quad g_1 = 0{\cdot}5361,$$

$$k_2 = 14{\cdot}3389, \quad g_2 = -0{\cdot}4156.$$

Entering our chart with these values of g_1 and g_2, we read off the estimates of ρ and δ as

$$r = 2{\cdot}70, \quad d = 3{\cdot}15.$$

Hence we find

$$s = 2{\cdot}202, \quad m_1 = 20{\cdot}238, \quad m_2 = 27{\cdot}175,$$

$$p_1 = r/(1+r) = 0{\cdot}730, \quad p_2 = 1 - p_1 = 0{\cdot}270.$$

The expected frequencies, based on the two normal components with a common standard deviation estimated by s, are shown in the third column of Table 3 and the differences (observed − theory) in the fourth. After combining four groups in each tail, it is found that $\chi^2 = 15{\cdot}28$, with 13 degrees of freedom. The probability of exceeding this value if our theoretical structure were correct is 0·29. The fit is quite good, but it is evident that there are certain systematic differences between theory and observation.

K. Pearson (1914, p. 131) did not assume a common standard deviation for the two components and used his general moment method of fitting (1894) involving the solution of a nonic. He found

$$m_1 = 19{\cdot}8926, \quad s_1 = 2{\cdot}0566; \qquad m_2 = 26{\cdot}2463, \quad s_2 = 2{\cdot}6260;$$

$$p_1 = 0{\cdot}6505, \quad p_2 = 0{\cdot}3495.$$

Using these estimates of the parameters, the theoretical distribution given in the fifth column of Table 3 is obtained, with the differences (observed − theory) of the sixth column. In this case $\chi^2 = 10{\cdot}77$ with 12 degrees of freedom, a value exceeded by chance with a probability of 0·55. The fit has been improved by allowing the two variances to differ, but there is still evidence of some systematic discrepancy, as shown by the run of the signs of differences.

6. SUMMARY

This graphical method of analysis into two components by means of the skewness-kurtosis diagram may be useful in many cases where a simple, rapid determination of the probable nature of the two sub-universes is wanted. It is less efficient and less accurate than existing theoretical methods, but there are always unavoidable errors due to sampling and necessary assumptions about the components. It is therefore likely to be of particular value, in contrast to these methods, in producing quick approximations to the components of small samples. Any kind of distribution may be analysed by this method into two normal components, irrespective of whether its character is, or is not, due to their physical existence, since all real points on the diagram lie within the bounding parabola; hence the method cannot offer any confirmation of their physical existence. It may, however, permit us to devise physical and experimental tests to identify their physical nature if their existence seems probable.

---

* Pearson gives values of moments, with n as divisor, not k-statistics, but with $n = 1000$ the slight adjustment called for seemed unnecessary.

Table 3

Central value	Observed frequency	Distribution fitted by diagram	Difference	Pearson's composite distribution	Difference
12	— } 10	0·2 } 11·4	−1·4	0·1 } 10·6	−0·6
13	1	0·6		0·5	
14	—	2·5		2·2	
15	9	8·1		7·8	
16	21	21·2	−0·2	21·6	−0·6
17	56	45·3	10·7	47·4	8·6
18	79	78·9	0·1	82·9	−3·9
19	114	112·3	1·7	115·1	−1·1
20	122	130·5	−8·5	128·0	−6·0
21	110	124·6	−14·6	115·8	−5·8
22	85	98·9	−13·9	89·2	−4·2
23	85	68·8	16·2	65·6	19·4
24	61	48·7	12·3	54·5	6·5
25	47	43·3	3·7	53·3	−6·3
26	49	46·7	2·3	54·3	−5·3
27	47	49·7	−2·7	51·0	−4·0
28	44	45·6	−1·6	42·4	1·6
29	31	34·8	−3·8	30·7	0·3
30	20	21·6	−1·6	19·2	0·8
31	11	11·0	0·0	10·5	0·5
32	4 } 8	4·6 } 6·7	1·3	4·9 } 7·9	0·1
33	4	1·6		2·0	
34	—	0·4		0·7	
35	—	0·1		0·1	
Total	1000	1000		1000	

I wish to record my sincere thanks to the Editors of *Biometrika* for their most generous and helpful advice and assistance in preparing the final draft of this paper; to my brother, Dr F. W. Preston, of Butler, Pa., U.S.A., through whose interest and at whose suggestion I originally began investigation of the subject; and to his wife Jane, and to Mr R. R. Lehnerd, of the Preston Laboratories, Butler, for their care and patience in making legible the original script and diagrams.

REFERENCES

Pearson, K. (1894). *Phil. Trans.* A, **185**, 71.
Pearson, K. (1914). *Biometrika*, **10**, 85.
Rao, C. R. (1948). *J.R. Statist. Soc.* B, **10**, 166.

A note on regions for tests of kurtosis

By G. E. P. BOX

Imperial Chemical Industries, Dyestuffs Division Headquarters, Blackley, Manchester 9

The most common test criterion for kurtosis is b_2, the sample fourth moment divided by the fourth power of the standard deviation; an alternative criterion proposed by Geary (1935) and denoted by a is the ratio of the sample mean deviation to the sample standard deviation. The present author has shown in a paper appearing on p. 318 of this volume that Bartlett's (1937) modification of the Neyman-Pearson L_1 criterion (1931) for testing for the equality of variances is so sensitive to kurtosis that when there is only one observation per group (group means assumed known) this criterion denoted by $M_{.1}$ is of the same order of sensitivity to kurtosis as is b_2 or a.

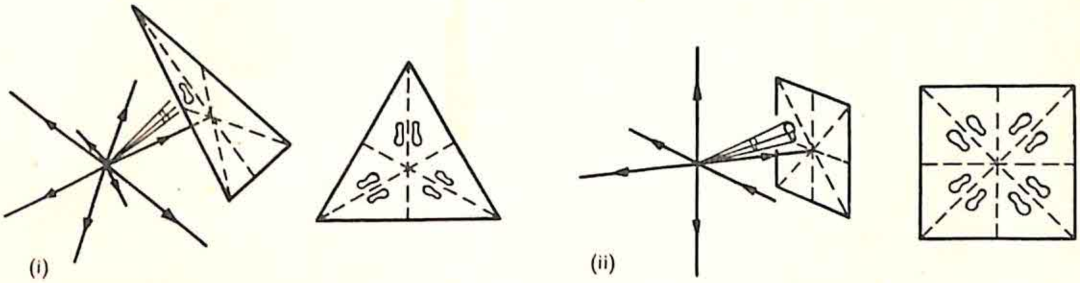
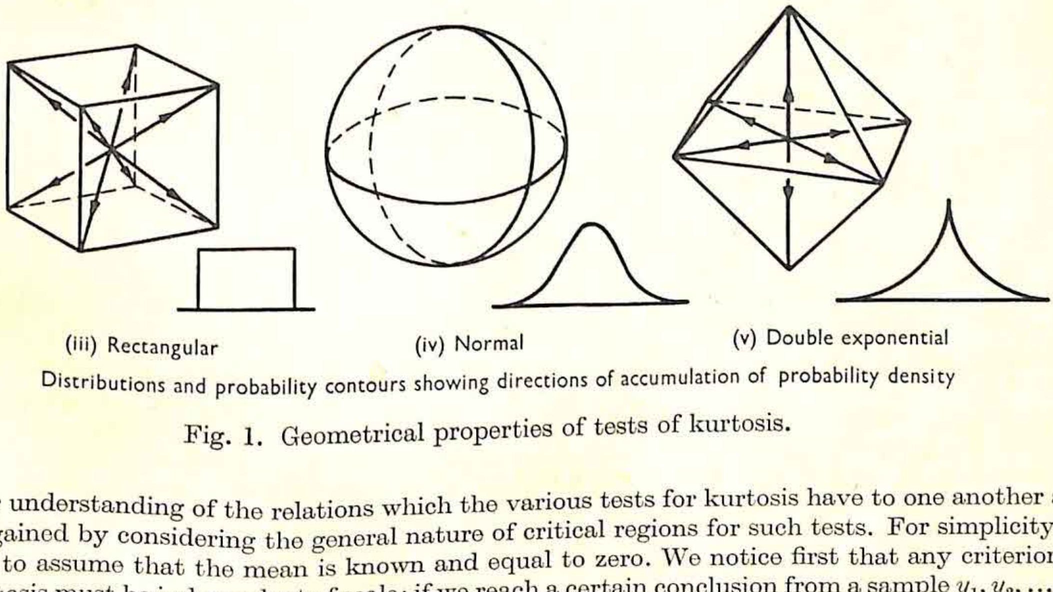


Illustration of symmetry conditions when $N = 3$



(iii) Rectangular (iv) Normal (v) Double exponential

Distributions and probability contours showing directions of accumulation of probability density

Fig. 1. Geometrical properties of tests of kurtosis.

A better understanding of the relations which the various tests for kurtosis have to one another and to $M_{.1}$ is gained by considering the general nature of critical regions for such tests. For simplicity we shall need to assume that the mean is known and equal to zero. We notice first that any criterion to detect kurtosis must be independent of scale; if we reach a certain conclusion from a sample $y_1, y_2, \ldots, y_N$ we must reach the same conclusion from the sample $ky_1, ky_2, \ldots, ky_N$. It follows that the boundary of the critical region in the sample space consists of one or more cones with apices at the origin. Also each observation has equal weight so that if the point $y_1, y_2, \ldots, y_N$ is on the boundary of a critical region then so must the remaining $N! - 1$ points obtained by permuting the order of the values. Finally, since a test of kurtosis should detect departures from normality other than those due to asymmetry of the distribution, the test criterion must be independent of the signs of the observations. Thus the pattern of critical subregions in the region of the space in which the observations are all positive will be repeated in all the remaining $2^N - 1$ regions of the space generated by possible differences in sign. The conditions of symmetry thus imposed are illustrated for $N = 3$ in Fig. 1 (i) and (ii) which shows cross-sections of subsets of cones satisfying these conditions. When the null-hypothesis of normality is true, the probability density in the sample space is constant over the surfaces of N-dimensional hyperspheres with centre at the origin. Consequently if the hypercones of the critical region are such that they include a proportion α of the total surface of such hyperspheres the error of the first kind is controlled at the level α irrespective of the value of the scale parameter.

In addition to controlling this error, we require the test to pick out as often as possible a departure from normal-kurtosis when it occurs. Consider the probability contours corresponding to symmetrical distributions having varying amounts of kurtosis in the case $N = 3$. For the normal distribution the contours are spheres (Fig. 1 (iv)). As γ_2 falls below zero, the contours tend to belly out and for the rectangular distribution ($\gamma_2 = -1{\cdot}2$) the single probability contour is the surface of a cube (Fig. 1 (iii)), the density within the cube being uniform. As γ_2 is made larger than zero the contours tend to sag inwards till for the double exponential distribution ($\gamma_2 = 3$) the contours are surfaces of regular octahedra (Fig. 1 (v)). In general for platykurtic distributions ($\gamma_2 < 0$) there is an accumulation of probability density along the equiangular lines (and a depletion along the co-ordinate axes), whilst for leptokurtic distributions ($\gamma > 0$) there is an accumulation of probability density along the co-ordinate axes (and a depletion in the direction of the equiangular lines). It follows that (i) a good critical region for testing for platykurtosis would be the 2^N hypercones having apices at the centre and axes along the 2^N equiangular lines, whilst (ii) a good critical region for testing for leptokurtosis would consist of the $2N$ hypercones having their apices along the $2N$ branches of the co-ordinate axes.

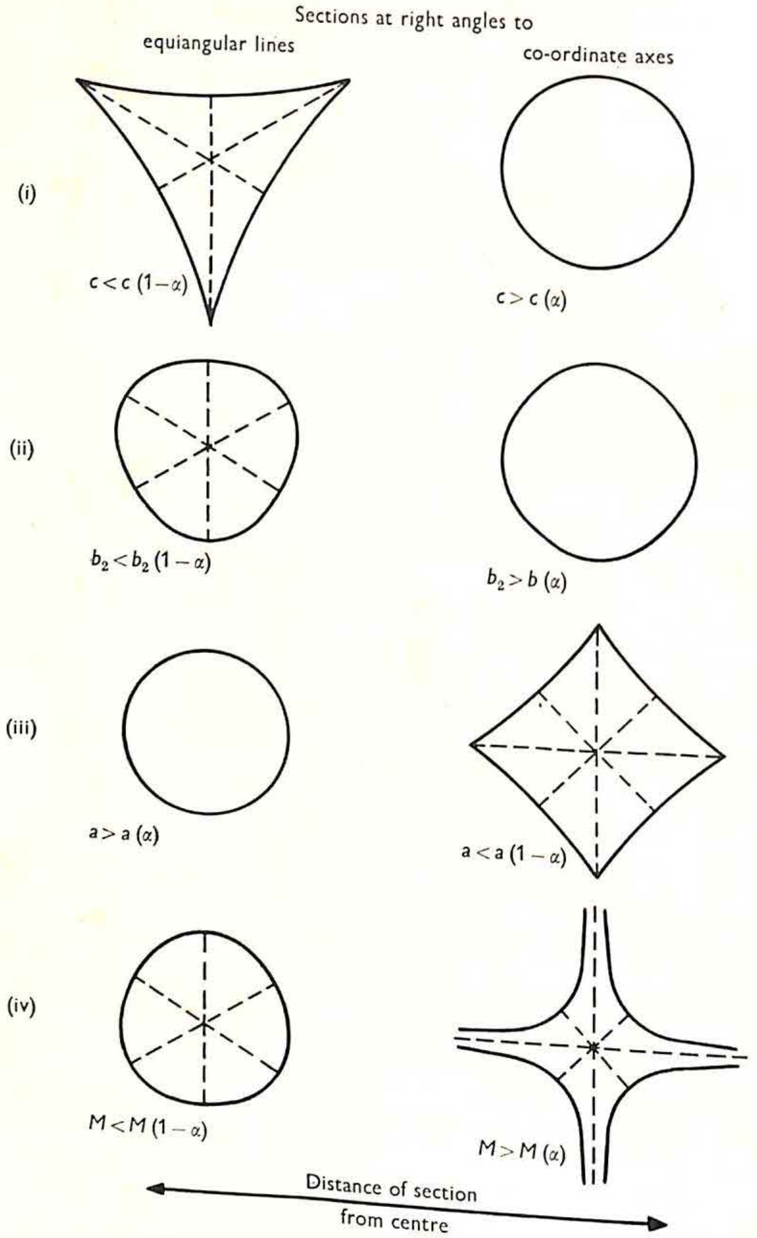


Fig. 2. Sections of critical regions.

The directions of the axes of cones forming such regions when $N = 3$ are shown by arrows in Fig. 1 (iii) and (v) respectively. In particular b_2 and a have critical regions of these types and so has $M_{.1}$. In each case, the particular type which occurs depends on whether we are using the criterion to test for leptokurtosis or for platykurtosis. The regions differ only in the shapes of their cross-section and even these differences are limited to some extent by the symmetry conditions listed above. For the case $N = 3$, sections of cones corresponding to the approximate upper and lower 5 % levels for the test criteria b_2, a and $M_{.1}$ are shown in Fig. 2 (ii), (iii) and (iv). (In order to make clear that the region in Fig. 2 (ii) defined by $b_2 > b_2(\alpha)$ has a non-circular section, it has been necessary slightly to exaggerate the departure from circularity.)

2. It is of some interest to consider kurtosis tests from the point of view of the Neyman-Pearson theory of best critical regions. Suppose that the distribution specified by the alternative hypothesis is a member of the family

$$p(y) = \left\{\frac{2\rho}{q}\Gamma\left(\frac{1}{q}\right)\right\}^{-1} \exp-\left\{\frac{|y|}{\rho}\right\}^{q} \quad (-\infty<y<+\infty, 0<q<\infty, 0<\rho<\infty), \tag{1}$$

and *q is given*. This is a symmetrical distribution with mean zero and

$$\kappa_2 = \rho^2\frac{\Gamma(3/q)}{\Gamma(1/q)}, \quad \gamma_2 = \frac{\Gamma(5/q)\,\Gamma(1/q)}{\{\Gamma(3/q)\}^2} - 3. \tag{2}$$

When $q = 2$, γ_2 is zero and the distribution is normal. When $q > 2$ the distribution is platykurtic and, in particular, when q tends to infinity the distribution tends to the rectangular, for (1) tends to

$$p(y)\begin{cases} = 1/(2\rho) & (y<\rho) \\ = 0 & (y>\rho). \end{cases} \tag{3}$$

When $q < 2$ the distributions are leptokurtic in particular when $q = 1$ the distribution is the double exponential.

The likelihood ratio test criterion (Neyman & Pearson, 1928) for testing the null hypothesis of normality ($q = 2$) against the *single alternative* that the distribution is of the form of the above equation with q equal to some other *specific* value q_0 is readily found to be proportional to

$$\Lambda_{q_0} = (m_{q_0})^{1/q_0}/s, \tag{4}$$

where m_q is the qth absolute moment,

$$m_q = \Sigma\,|\,y\,|^q/N, \quad \text{and in particular } s = m_2^{\frac{1}{2}}. \tag{5}$$

The inequality

$$\Lambda_{q_0} < \Lambda_{q_0}(1-\alpha), \tag{6}$$

where $\Lambda(1-\alpha)$ and $\Lambda(\alpha)$ refer respectively to the lower and upper significance points of the criterion, defines a set of cones lying along the co-ordinate axes when $q_0 < 2$ and a set of cones lying along the equiangular lines when $q_0 > 2$. That every such set defines a *best* critical region for the appropriate value of q_0 can be seen from the fact that on every* hyperspherical shell (4) and (6) define the region, independent of scale factors in the null distribution or in the alternative distribution, the boundary of which is given by the fundamental inequality $p_t \geqslant kp_0$ of Neyman & Pearson (1933). This region on every such shell thus contains for its size the greatest possible concentration of probability density when the appropriate alternative hypothesis is true.

We see that, $a < a(1-\alpha)$ defines the most powerful test when the alternative distribution is the double exponential, for which $\gamma_2 = 3$, also $b_2 < b_2(1-\alpha)$ defines the most powerful test when the alternative distribution is $p(y) = \text{const.}\exp-(y/\rho)^4$ for which $\gamma_2 = -0{\cdot}812$. When the alternative distribution is rectangular ($\gamma_2 = -1{\cdot}2, q = \infty$) the test is defined by $c < c(1-\alpha)$, where

$$c = |\,y_L\,|/s \tag{7}$$

and y_L is the observation largest in absolute magnitude. It should be noted that with this criterion we should be testing whether the largest deviate in absolute magnitude were *too small* and the test would be based on the *lower* tail area of the distribution of c. The same criterion, but using the upper tail area of

---

* This is literally true if it is assumed as in (1) that $q_0 < \infty$. In the limiting case when $q_0 = \infty$, if we suppose a *particular* rectangular distribution of semi-range ρ to be the alternative and if r is the radius of a hemispherical shell, then the region defined on the shell will contain an amount of probability density *greater* than any other region of similar size if $\rho < r < \sqrt{N}\rho$. For all other values of r it will contain an amount of probability density *equal* to that of other regions of the same size. The test is thus still most powerful in this limiting case.

c, has been proposed by Thompson (1935) as a test for a single outlier (see also Pearson & Chandra Sekar, 1936). Sections of critical regions for this criterion are shown in Fig. 2 (i). It will be noted that only the lower tail areas of c, b_2 and a define most powerful tests for the alternatives considered. In Fig. 2 the regions are differentiated by drawing in the axes of symmetry only for those sections which correspond to most powerful tests. We should expect c and b_2 to be particularly good criteria, therefore, when the alternative hypothesis was that the distribution showed marked platykurtosis; on the other hand, in testing for marked leptokurtosis a would be expected to be better. Pearson's (1935) sampling experiments (his Table 7) which are shown again in table 4 of the article appearing on p. 318 of this journal support this suggestion (although with so few results no firm conclusions can be drawn). For the rectangular distribution a larger number of significant results is obtained with b_2, whilst for the double exponential a larger number is obtained with a. It does not of course follow that a would become better than b_2 as soon as γ_2 was greater than zero; the change-over might well be at some other point.

In a study of possible tests of kurtosis of this type, and assuming the alternative distribution to be the symmetrical Gram-Charlier, Geary (1947) reached the conclusion that b_2 was the most 'efficient' test for *infinitely large samples* (and correspondingly small departures from normality). He also showed that, over a wide range of values of q, the power of the various possible tests would not be expected to be greatly different. The present research emphasizes that the choice of criterion for samples of finite size must depend on the type of alternative hypothesis which is in mind. The similarity of the critical regions for the various criteria confirm Geary's second conclusion.

REFERENCES

Bartlett, M. S. (1937). *Proc. Roy. Soc.* A, **160**, 268.
Geary, R. C. (1935). *Biometrika*, **27**, 310.
Geary, R. C. (1947). *Biometrika*, **34**, 209.
Neyman, J. & Pearson, E. S. (1928). *Biometrika*, **20** A, 175.
Neyman, J. & Pearson, E. S. (1933). *Phil. Trans.* A, **231**, 289.
Pearson, E. S. (1935). *Biometrika*, **27**, 333.
Pearson, E. S. & Chandra Sekar, C. (1936). *Biometrika*, 28, 308.
Thompson, W. R. (1935). *Ann. Math. Statist.* **6**, 214.

The frequency justification of sequential tests—addendum

By G. A. BARNARD, *Imperial College*

In my paper (1952) on the above topic, at the end of §3 (p. 147) an inversion of limit operations was performed without explicit justification. This has caused difficulty for some readers, and a justification for this process is now appended.

The omission of this proof in the first place was quite deliberate. In the opinion of the writer, statistics is a branch of applied mathematics in which undue attention to pure rigour is out of place. In other branches of applied mathematics, in mechanics, for example, one is expected to make use of experience to fill in what would otherwise be gaps in proofs. If a plane lamina is under consideration, and its surface density is given, one does not usually find the assumption explicitly mentioned that the lamina is a measurable set of points before use is made of the fact that the lamina will have a determinate mass. But in certain quarters great emphasis seems to be laid on just such assumptions in statistics.

It is not that the present writer is opposed to purely mathematical arguments. It is, that such arguments should be kept in their proper place, and that we should not pretend we are doing statistics when we are really doing nothing but pure mathematics.

In §5 of the paper, it is stated that 'Jeffreys has not ... laid down general principles which determine the choice of prior distributions in cases of "ignorance"'. This should read '...which *uniquely* determine the choice...'.

We now give the proof referred to above.

We wish to show that if

$$(a)\quad (1/\log b/a)\int_a^b \alpha''(\sigma,a,b)\,d\sigma/\sigma = \alpha' \quad \text{for all } 0<a<b<\infty$$

and

$$(b)\quad \text{as } a\to 0 \text{ and } b\to\infty,\ \alpha''(\sigma,a,b)\to\alpha,$$

then $\alpha = \alpha'$.

We notice first that $\alpha''(\sigma, a, b) = \alpha''(c\sigma, a/c, b/c)$ for any positive c. For if we make the changes $X \to cX$, $a \to a/c$, $b \to b/c$, then $\sigma \to c\sigma$, $s \to cs$, sa, sb, and t remain unaltered, and so $\lambda_n(a, b)$ remains unaltered. It follows that we can write $\alpha''(\sigma, a, b) = \alpha''(\sigma a, \sigma b)$.

We now put $t = \log \sigma$, $l_1 = \log a$, $l_2 = \log b$, in the integral, and put $\alpha''(\sigma a, \sigma b) = f(t + l_1, t + l_2)$. Then ($a$) becomes

$$(1/(l_2 - l_1)) \int_{l_1}^{l_2} f(t + l_1, t + l_2)\, dt = \alpha' \text{ for all } l_1 < l_2,$$

while (b) gives $f(x, y) \to \alpha$ as $x \to -\infty$ and $y \to +\infty$. The second condition implies that, given any positive ϵ, we can find an A such that

$$|f(x, y) - \alpha| < \epsilon \text{ whenever } x < -A \text{ and } y > A.$$

Now in (a), put $l_1 = -nA$, $l_2 = +nA$, with $n > 1$. Then we have

$$\begin{aligned}\alpha' &= (1/2nA)\left\{\int_{-nA}^{-(n-1)A} f(t - nA, t + nA)\, dt + \int_{-(n-1)A}^{+(n-1)A} f(t - nA, t + nA)\, dt + \int_{(n-1)A}^{nA} f(t - nA, t + nA)\, dt\right\} \\ &= (1/2nA)\{Af(t_1 - nA, t_1 + nA) + 2(n-1)Af(t_2 - nA, t_2 + nA) + Af(t_3 - nA, t_3 + nA)\},\end{aligned}$$

where $\quad -nA < t_1 < -(n-1)A, \quad -(n-1)A < t_2 < (n-1)A \quad \text{and} \quad (n-1)A < t_3 < nA,$

by the mean-value theorem. Now f is a probability, and is therefore bounded, so that as $n \to \infty$, the first and last terms within the braces remain bounded. Hence, by taking n large enough, we can make α' differ by less than ϵ from $f(t_2 - nA, t_2 + nA)$. But $t_2 - nA < -A$ and $t_2 + nA > A$, so that $f(t_2 - nA, t_2 + nA)$ differs by less than ϵ from α. Hence α and α' differ by less than 2ϵ.

REFERENCE

Barnard, G. A. (1952). The frequency justification of certain sequential tests. *Biometrika*, **39**, 144.

REVIEWS

The Design and Analysis of Experiments. By O. KEMPTHORNE. New York: John Wiley and Sons; London: Chapman and Hall. 1952. Pp. xix + 631. 76*s.*

Until two or three years ago there were few connected and comprehensive accounts of the theory and practice of experimental design. There were the classical monographs by Fisher and Yates and a vast journal literature on subsequent developments. To the statistician outside the field of agricultural experimentation, the ingenuity shown in devising more and more refined and specialized designs was indeed impressive, but the resulting complexity became, it must be admitted, somewhat baffling.

There was clearly a need for a book which would give a unified account of all the designs, their advantages and disadvantages and their relation to general experimental principles, and which would connect all this with modern statistical theory. As is inevitable when a gap in the literature is disclosed, not one but several books were written to fill it. Quite apart from those recent statistical text-books which have given prominence to experimental design (among which that by Bancroft and Anderson is a notable example) at least four books entirely devoted to this subject have appeared in the last four years. (It is regrettable, incidentally, that their titles are virtually identical.) The first, by Mann (*Analysis and Design of Experiments*), concerns itself strictly with the mathematics underlying the analysis of different designs. The second is the well-known volume by Cochran and Cox. The third, by Kempthorne, is the one under review and the fourth, by Quenouille (*The Design and Analysis of Experiment*), only appeared in the United Kingdom in early 1953. It is notable, though in no way surprising, that three of these books are written (or partly written) by statisticians who have worked at Rothamsted.

As might be expected with two books from the same publisher, the books by Cochran and Cox and by Kempthorne differ in approach. Professor Kempthorne says that 'the basic requirements of the user of experimental designs have been satisfied by the book of Cochran and Cox....The aim of this book has been to give a description of the design of experiments from as broad a view as possible and to relate the subject matter of this field of statistics to the general theory of statistics and to the general problem of experimental inference.' This should not be taken to mean that Cochran and Cox offer no guidance on general principles of design or on the advantages and otherwise of different designs, for they are useful and clear on both; but their book is a manual of experimental design, that by Kempthorne is an advanced text-book. It is the former which will be of greater service to most practical experimentalists; the latter which, by and large, will be of more interest to the theoretical statistician, whether as consultant or teacher. There is something to be said for using the two books in conjunction. Possibly because its approach is more difficult, Kempthorne's text does not always attain the clarity of exposition which distinguishes Cochran and Cox's book, and reference from one to the other often helps. In any case, this new book is not for the beginner, nor is it suitable for superficial and easy reference. It should be looked upon as a comprehensive and advanced text-book and studied with the concentration and patience such a book deserves.

Its comprehensiveness is indeed remarkable and Professor Kempthorne must be congratulated for carrying through so successfully the enormous task of providing a systematic treatment of experimental designs. He begins with six more or less introductory chapters, the first two on experimental principles, the next on elementary statistical notions (the value of this chapter is open to doubt; it is too scanty for those without sufficient background and superfluous for others) and three on least squares and general linear hypothesis theory. Chapters 7 and 8 between them give an admirable treatment of randomization. Chapters 9 and 10 deal with the two basic designs, randomized blocks and latin squares, and are followed by one on plot technique. Chapter 12 is on the sensitivity of experiments, a subject which has had relatively little attention in other texts. Then follows the core of the book, namely, nine chapters dealing with factorial designs. After this come four chapters on lattice designs, two on incomplete block designs, one on the analysis of groups of experiments and a final one, chapter 29, on treatments applied in sequence. The mathematical theory is given for all the designs, together with explanations and numerical examples. Judged by any standard, the book is a fine achievement.

As there will undoubtedly be future editions, it may be worth mentioning one general aspect which could be improved. Of some subjects treated by Professor Kempthorne one has the impression, not so much of a unified treatment, as of a series of notes scattered throughout the volume. The analysis of covariance is a case in point. It is first introduced in § 6·6 as a modification to the general two-way model; the mathematics for it are given at length, but no explanation is given of its purpose or use. This comes later, first in § 8·7 in the chapter on The Validity of Analyses of Randomized Experiments and then in § 9·9 in the chapter on Randomized Blocks. (The index does not refer the reader to § 8·7). A certain amount of rearrangement on this, as on some other topics, would clarify the presentation. The very first opportunity should also be taken for improving the index and the cross-referencing. The present index is quite inadequate and there is virtually no cross-referencing in the text.

The theory and practice of experimental design has grown so much out of agricultural experimentation that it is inevitable that the discussion and illustrations in this book refer almost wholly to this field. At several points, Professor Kempthorne remarks critically on methods used in the social sciences. It is of course true that experimentation in this field is very much more difficult than in agriculture. But it is not impossible nor are survey techniques as limited in their value as Professor Kempthorne suggests, in passing, in his first chapter. In the last year or two, attempts have been made to use experimental designs in investigations dealing with human populations and something has been learned about the difficulties and possibilities. For the statistician working in the social sciences, one of the most interesting problems is, to what extent experimental techniques, such as those discussed in this book, can be imported into his domain of study. For him, as for his statistical colleagues in other sciences and fields of activity, Professor Kempthorne's book is a most welcome addition to the literature.

C. A. MOSER
London School of Economics

Associated Measurements. By M. H. QUENOUILLE. Pp. x + 242. Butterworth. 35*s*.

The Design and Analysis of Experiment. By M. H. QUENOUILLE. Pp. xiii + 356. Griffin. 36*s*.

To have produced these two books within a comparatively short time is a notable achievement. The books have much in common; they are advanced non-mathematical text-books on statistical methods excellently illustrated with examples, mostly biological, which are handled with great skill and with a clear sense of the practical purpose of the statistical analysis. The main fault is an occasional carelessness of writing.

The discussion is mostly in terms of specific examples and, while the account of these is remarkably clear, the reader without a sound background of theory may find it difficult to apply the more advanced methods to cases slightly different from those considered by Mr Quenouille. For example, the method of least squares, which underlies much of both books, is, so far as I could see, never mentioned and is certainly never explained. It is true that the stage-by-stage fitting of groups of constants in a least-squares model is often most expeditiously done by analysis of covariance, and that this technique is used repeatedly. But a statement in some of the more complicated cases of the form that is assumed for the underlying population would have given a unity which is at present lacking. An excellent point about both books is the care taken to explain the assumptions necessary for the validity of the various methods.

Associated Measurements starts with a very good section on graphical analysis, including an account of non-parametric tests for association. The description here of tests of partial correlation is open to criticism. The author defines a *coefficient of medial correlation* between two variates x and y by dichotomizing the scatter diagram at the medians, finding the proportion, p, of observations in the positive quadrant, and putting

$$\phi_{xy} = 4p - 1. \qquad (1)$$

For an infinite population ϕ_{xy} takes values plus and minus unity for perfect correlation, and zero for independence. Also when x and y are independent, the distribution of ϕ_{xy} does not depend on the form of the population, so that a non-parametric test of independence may be obtained. For a bivariate normal population ϕ_{xy} is related to the product-moment correlation coefficient, ρ_{xy}, by W. F. Sheppard's formula

$$\rho_{xy} = \sin\left(\frac{\pi\phi_{xy}}{2}\right). \qquad (2)$$

Mr Quenouille next defines a partial medial correlation coefficient by

$$\phi_{xy.z} = \frac{\phi_{xy} - \phi_{xz}\phi_{yz}}{(1 - \phi_{xz}^2\phi_{yz}^2)^{\frac{1}{2}}}. \qquad (3)$$

No derivation of this formula is given, and although it is used consistently, one suspects that it is a misprint for

$$\phi'_{xy.z} = \frac{\phi_{xy}-\phi_{xz}\phi_{yz}}{(1-\phi^2_{xz})^{\frac{1}{2}}\,(1-\phi^2_{yz})^{\frac{1}{2}}}, \qquad (4)$$

derived by analogy with product-moment correlations. Both (3) and (4) can give quite misleading results. To see this it is enough to consider some particular trivariate normal populations. For example, suppose that $x=y+z$, where y and z are independently distributed in unit normal distributions. Then, conditionally on z, x is perfectly correlated with y. But $\phi_{xy.z}=\phi_{xy}=\frac{1}{2}$ and $\phi'_{xy.z}=1/\sqrt{3}$, so that the partial ϕ's suggest that the correlation is unchanged, or only slightly improved, by holding z fixed. The situation is more satisfactory if, conditionally on z, x is independent of y. For, in this case, $\phi_{xy.z}$ and $\phi'_{xy.z}$, while not necessarily zero, are always appreciably smaller than ϕ_{xy}. However the vanishing of $\phi_{xy.z}$ and $\phi'_{xy.z}$ does not imply that $\rho_{xy.z}$ is small. For example if $\rho_{xy}=0{\cdot}696$, $\rho_{xz}=\rho_{yz}=0{\cdot}891$, then $\rho_{xy.z}=-0{\cdot}476$, although $\phi_{xy.z}=\phi'_{xy.z}=0$. Also by taking ρ_{xy}, etc. near unity, $\rho_{xy.z}$ may be made arbitrarily near minus one, with $\phi_{xy.z}=\phi'_{xy.z}=0$. Enough has been said to make it clear that the uncritical use of partial ϕ could be very misleading.

Similar remarks may, incidentally, be made about the coefficient of partial rank correlation given by Prof. M. G. Kendall in his *Rank Correlation Methods*. Although the theory of rankings derived by sampling an underlying normal population is only a very special case of the general theory of ranking, it is worth noting that in such cases Kendall's partial τ may give a wrong idea of the form of the continuous population. The similarity with ϕ arises because $\bar{\tau}$, the expectation of τ in samples from a normal universe, is given by the analogue of (2),

$$\rho=\sin\left(\frac{\pi\bar{\tau}}{2}\right). \qquad (5)$$

When the underlying population is normal these difficulties may of course be avoided by transforming ϕ or τ by (2) or (5) before calculating the partial correlation. There are a number of objections to this as a general procedure.

The second section of *Associated Measurements* is called Numerical Analysis and contains a brief account of analysis of variance, analysis of covariance and regression analysis. The next section, on Rapid Estimation and Analysis, includes chapters on the quick estimation of regressions by the coarse grouping of the observations, and on further uses of analysis of covariance. The last three chapters are entitled Analytical Complications, Time Series Analysis and Multivariate Analysis. The chapter on time series deals mainly with useful approximate methods based on correlation coefficients and analysis of covariance. The final chapter on multivariate analysis seems to me one of the best brief introductions to the subject at present available.

The Design and Analysis of Experiment begins with a twenty-page account of the general principles of design. This is so terse that it is unlikely to make much appeal to an experimenter not already fairly familiar with the methods described. For example there is a bald statement that randomization is essential if unbiased estimates and unbiased estimates of error are to be obtained, but there is no attempt to explain in non-statistical language just what the consequences are of failure to randomize.

The following chapters deal with randomized blocks, Latin squares, and with simple and general factorial and split-plot designs. The emphasis tends to be on the appropriate forms of analysis of variance and covariance.

Section B consists of four chapters on incomplete block designs. The main types of confounding and fractional replication are well described. The advice (p. 111) that, except in special cases, the interactions to be confounded should be chosen at random from interactions of a given order, is surprising. Such a randomization is in any case quite pointless, and is dangerous in that it tends to distract attention from the great importance of a careful choice of the interactions to be confounded. The fairly general condemnation of partial confounding (p. 117), because of the complexity of the calculations involved in the analysis, is also open to question.

Sections C and D entitled Long-term Policy and Experimental Complications contain a good deal of interesting material, the discussion of the combination of results from a series of experiments being particularly good.

To sum up, while these books should be read critically, they are very useful and should appeal to anyone wanting a non-mathematical account of advanced statistical methods.

D. R. COX

The Statistics of Bio-assay. By C. I. BLISS. New York: Academic Press, Inc. Pp. iii + 184. $3.50.

Statistical Method in Biological Assay. By D. J. FINNEY. London: Charles Griffin and Co. Ltd. Pp. xix + 661. 68*s*.

Probit Analysis. By D. J. FINNEY. Cambridge University Press. 2nd ed., pp. xiv + 318. 35*s*.

The statistical theory of bio-assay has been almost entirely developed during the past 20 years and has reached a reasonably stable state, so that the appearance of text-books by two of the foremost workers in this field is extremely welcome. The two general texts cover much the same ground, though Finney's book is by far the more detailed; Bliss's book originally appeared as a single chapter in György's *Vitamin Methods* and is necessarily rather compressed. Both authors deal with parallel line assays with quantitative and quantal responses, slope-ratio assays, the use of covariance and the combination of potency estimates; Bliss has an interesting section on control charts, while Finney includes several topics which do not arise in vitamin work and so fall outside the scope of the shorter book.

Inevitably in so brief a compass, Bliss's book is little more than a collection of recipes—indeed he refers his readers to Finney for all mathematical details. Within this limitation it should prove very valuable. The needs of the practising biologist are always kept in mind, and the author is bold enough to suggest approximate rule-of-thumb techniques for important problems, rather than to decline to commit himself on the grounds that no exact solution is yet available.

Finney's book is substantial to the point of being encyclopaedic, and some of the theory set out is well ahead of practice, at any rate in the assay field. The reader is supposed to be reasonably familiar with the basic statistical techniques, but apart from this, the suggested analytical methods are explained in full detail with an excellent set of worked examples. The underlying logic of bio-assay is discussed at length, and a number of incomplete block designs are given which should prove of great practical use. Four chapters are devoted to quantal assays, and these are noteworthy for a very full discussion of the various alternatives to probit analysis that have been suggested from time to time. The book includes a 40 page Appendix of Tables, many of them new.

The choice between these two books is not an easy one. Statisticians will probably prefer Finney. Biologists may like to keep Bliss's book on the bench (not least because of its handy size), but a study of Finney's book will be necessary for non-routine problems, and may well enable them to improve on their accepted techniques.

A second edition of Finney's *Probit Analysis* is very welcome, the first edition having been unobtainable for some time. Few changes have been made in the original material, but a substantial new chapter on Recent Developments has been added, and some of the tabular matter is new (the table of weighting coefficients now covers natural mortality up to 90 %!). Alternatives to probit analysis are still rather inadequately treated (as the author himself points out), but this has been remedied by the very full discussion in the later book. The omission of one or two other topics, such as Tocher's analysis of grouped data and Finney's own work on graphical methods and the adequacy of a single cycle of calculation, seems to be due to a long delay in publication—the preface to the new edition is dated December 1949. The standard of production leaves something to be desired; the new edition has been produced by an offset-litho process on rather thick paper, and the appearance of the page has suffered somewhat. It seems a pity that the price should still be almost double that of the original edition. Nevertheless, *Probit Analysis* remains an essential book for the many biologists and statisticians who work with sigmoid curves.

M. J. R. HEALY

Statistical Theory with Engineering Applications. By A. HALD. New York: Wiley; London: Chapman and Hall. 1952. Pp. xii + 783. 72*s*.

Statistical Tables and Formulas. By A. HALD. New York: Wiley; London: Chapman and Hall. 1952. Pp. 97. 20*s*.

This book has as its aim the provision, in simple and logical form, of those parts of statistical methods which are of use to the engineer in his daily work. Nevertheless, in its 800 pages it covers considerably more ground, and that ground in greater detail, than might be expected from this simple aim. The result has been to produce a text-book which derives from first principles and illustrates with practical examples all the common statistical processes.

The book assumes no previous knowledge, and although advanced mathematical techniques are avoided the treatment and mode of development is mathematical in form. Nearly all the important results derived mathematically are subsequently illustrated practically. The contents include probability, graphs, the fundamental distributions, the analysis of variance, sampling investigations, the analysis of regressions and sequential analysis. It is noteworthy that the various distributions which are studied take up nearly 350 pages, which gives some idea of the thoroughness of the treatment. The many practical examples that are provided to illustrate the theory are drawn almost exclusively from engineering sources but they are put down in such a way that no detailed knowledge of engineering technique is required and the reader could easily substitute his own examples from his particular sphere of applications. The only criticism that could be levelled against the examples is that they are all chosen to illustrate just one point at a time, and one feels that, at any rate towards the end of the book, more complex illustrations should be used bringing in simultaneously a number of points from the theoretical development. As it is, few practical problems that the engineer is likely to meet with will be as simple as those given in the book. Also given in conjunction with each piece of theory are references to original memoirs and papers which describe practical illustrations. The reviewer feels that such references, instead of being scattered, would be better gathered together at the end of each chapter. One other minor point is that far too many equations are numbered, thereby giving them a certain prestige value. For example, in Chapter 18 no less than 157 equations receive numbers and few are referred to again nor are they likely to be quoted.

The particular portion of the book that struck the reviewer as one of the best of its kind was the section on the analysis of variance which is developed extremely logically. The hierarchical and cross-classification types of problem are discussed in some detail from the view-point of both the systematic and the random models with a very lucid explanation of the concept of an interaction. Later the case where there is a mixture of the two models is discussed. The technique of the analysis of variance has great value for the engineer and it is rather a pity that the treatment here is not quite as full as in other parts of the book. Two examples of this may suffice. The analysis of variance is built up on the twin assumptions that the residuals have equal variances and are normally distributed. If the former does not hold then a transformation can sometimes be applied and although this is mentioned on page 427 the full implications are not brought home. The latter form of departure would seem at first sight to indicate that the usual F or z test will not give the true significance level. It has, however, been shown that the non-normality of the residuals does not, in fact, affect the theoretical significance levels very much. Nevertheless if this point is not discussed it is inclined to worry the student. Secondly a short account of the simpler techniques of probit analysis would have been an extremely useful addition to the chapter on linear regression. Perhaps also, in a book of this size, a fuller account of time series might have been expected.

All authors have their own ideas as to the logical order of development of their subject. In this book the discontinuous distributions such as the binomial are dealt with in outline only before the main part of the book on continuous distributions and they are then returned to in more detail later. An alternative scheme, which would appear to have much in its favour, would have been to have dealt with all distributions in this manner and then to devote later chapters to problems of estimation, confidence limits, testing hypotheses and so on, bringing in the various distributions already discussed.

As a companion to the book a paper-covered set of tables has been issued and there are many references in one volume to the other. These tables are partly statistical, such as those for the normal and t distributions, and those giving confidence limits for the binomial probability, and partly mathematical, such as those for $\log n!$, square roots and reciprocals. All told there are nineteen tables and they cover the work in the book. Viewed in this light they are adequate and are set out in the most useful manner once the reader has got used to the word 'fractile' for significance level. They are, however, not full enough to stand by themselves as a book of statistical tables, especially since the introduction provided is only of very limited value without the book at hand as a guide.

Summarizing, one may classify this book of Hald's as a text-book, pure and simple. There are no discursions into the realms of debatable or controversial issues and hence it is a useful work which can be studied with much profit by a person intent on getting a sound working knowledge of statistics whether engineer or no. At the same time there is a risk that its length may put off all but the keenest student, particularly if he must find the money to buy it himself.

P. G. MOORE

Statistical Decision Functions. By A. Wald. New York: John Wiley and Sons, Inc.; London: Chapman and Hall, Ltd. 1950. Pp. ix + 179. 44*s*.

By the tragic death of Abraham Wald in December 1950 the world of mathematical statistics lost one of its most active and ingenious minds. The present book, published just before his death, shows him to have been at the height of his powers.

To explain what a decision function is, and to introduce concretely the leading ideas defined by Wald, we will consider the following simple example:

An urn contains black balls and white balls, and the proportion p of black balls is unknown. A fixed number n of balls is drawn, at random with replacement, and r of them are observed to be black. We are required to make a guess, p', at the value of p, and if we are wrong we lose an amount of money proportional to $L=(p-p')^2$.

Any rule by means of which we arrive at our guess will be a decision function for this situation. Since n is fixed, the only quantity on which our guess can depend is r, but the value of r need not uniquely determine our guess; we may make p' depend also on the toss of a coin or some other random event. Thus examples of decision functions $d(r)$ in this situation are:

$$\begin{aligned}&\text{(i)}\ \ d(r)=r/n,\quad \text{(ii)}\ \ d(r)=\tfrac{1}{2},\\ &\text{(iii)}\ \ d(r)=\lambda(r/n)+\mu(\tfrac{1}{2}),\ \text{for some}\ \lambda=1-\mu,\ 0<\lambda<1,\\ &\text{(iv)}\ \ d(r)=r/n\ \text{with probability}\ \tfrac{1}{2}\\ &\qquad\quad\ =\tfrac{1}{2}\ \text{with probability}\ \tfrac{1}{2}.\end{aligned}$$

In cases (i)–(iii) we take $p'=d(r)$, while in case (iv), which is an example of a *randomized* decision function, we toss a coin, and if it gives heads we take $p'=r/n$, while if it gives tails we take $p'=\frac{1}{2}$.

If we could attach, to each decision function d, a definite monetary loss incurred by its use, we would find the problem of choosing between one decision function and another quite simple. We would simply choose that function which minimized the loss. But such a simple solution is impossible, for two reasons. For one thing, for a given value of p the value of r will vary randomly, and so the guess p' will vary, and the loss L will vary along with p'; and for another thing, the distribution of p' will in general depend on p, and, for a given p', the loss L depends on p. Wald disposes of the first difficulty by noting that, for given p, and given d, the distribution of p', and hence that of L, is fixed, so that we may evaluate the mean value of L as a function of p and d. This function Wald calls the risk function, $\rho(p, d)$.

In the four examples we have given the values of ρ are

$$\text{(i)}\ pq/n,\ \text{(ii)}\ \tfrac{1}{4}-pq,\ \text{(iii)}\ \lambda^2pq/n+\mu^2(\tfrac{1}{4}-pq),\ \text{(iv)}\ \tfrac{1}{8}-pq(1-(1/n))/2.$$

Wald then adopts the principle that it is reasonable to act so as to minimize the expected value of the loss, when this can be calculated. This implies that if we have two risk functions, $\rho(p, d_1)$, $\rho(p, d_2)$, the first of which is, for a given value of p, less than the second, then, for this value of p, the first decision function is 'better' than the second, in the sense that it is reasonable to act on the first decision function rather than the second. If the first decision function is sometimes better, and is never worse, than the second, Wald says the first is 'uniformly better' than the second. Taking our decision functions (i)–(iv), we see that (i) is best of all when $p=0$ or $p=1$, since then its risk function is zero; while (ii) is best of all when $p=\frac{1}{2}$; on the other hand, when $p=\frac{1}{2}$, (i) is worst of all, while when $p=0$ or 1, (ii) is worst of all. If we from now on take $\lambda=\sqrt{n}/(1+\sqrt{n})$, the risk function (iii) becomes $\frac{1}{4}(1+\sqrt{n})^2$, which is independent of p and uniformly less than (iv). Thus (iii) is uniformly better than (iv), but it is not uniformly better than (i) or (ii); nor is either (i) or (ii) uniformly better than (iii). Thus, if we confine attention to these four decision functions, the principle of minimizing the mean loss enables us to rule out (iv) from consideration, but leaves us with (i), (ii) and (iii) to choose from. These three are what Wald calls 'admissible' decision functions, because for none of them does there exist another decision function which is uniformly better. (iv) is not admissible. If Mr A uses an admissible decision function to guide his conduct, no one who accepts the principle of minimizing mean loss can say that Mr A is acting unreasonably, since for some possible situations the rule Mr A is using is at least as good as any other.

We can go further if we are prepared to rule out cases where it is impossible to minimize the mean loss—such as would arise in the example we have taken if we allowed n to vary and assumed that further observations cost nothing. In this latter case, for any rule leading to a decision, we could clearly get a better one by taking more observations. But if we rule out such cases, then accepting the principle of minimizing mean loss implies that we need consider as reasonable only admissible decision functions. Thus the problem of determining the class of all admissible decision functions is seen to be of central importance.

Unfortunately, Wald was unable to formulate simple and general conditions which guarantee that exceptional cases having the same effect as the 'free observations' of the previous paragraph are ruled out. He therefore introduces, instead of the class of all admissible decision functions, the notion of a 'complete class' of decision functions. A class of decision functions is said to be complete if for any decision function d not in the class there is a decision function d' in the class which is uniformly better than d. If we confine our attention to decision functions belonging to a complete class we shall not lose anything thereby, since those we are ignoring are worse than some we are considering. The set of all possible decision functions is itself a complete class; so that the idea of a complete class helps us to choose our decision function only in so far as we make our complete class as small as possible. Wald obtains here what is probably the best possible result, viz. that under very general conditions, the set of all Bayes solutions forms a complete class of decision functions.

By a Bayes solution of a decision problem is meant a solution arrived at by giving the unknown parameter (p in our case) a prior distribution, and then using Bayes's theorem to minimize the mean loss. For example, our decision function (iii) can be arrived at by giving p the prior probability function $Kp^{\gamma-1}q^{\gamma-1}$, with $\gamma=\sqrt{n}/2$. When r black balls out of n have been observed, the posterior distribution of p is then given by the probability function

$$Kp^r q^s p^{\gamma-1} q^{\gamma-1} \Big/ \int_0^1 Kp^r q^s p^{\gamma-1} q^{\gamma-1} dp,$$

and to minimize the mean squared error we take the mean of this distribution, $(r+\gamma)/(n+2\gamma)$, which with the given value of γ is the same as (iii). The decision function (ii) is a Bayes solution, since it can be obtained by taking it as *a priori* certain that $p=\frac{1}{2}$. On the other hand, the decision function (iv) cannot be obtained as a Bayes solution since we shall always take the mean of the posterior distribution, and this will always be a definite number, not a random variable. Finally, (i) can be arrived at as a limit of Bayes solutions corresponding to the prior distributions given by K/pq for $\alpha<p<1-\alpha$, as α tends to 0. A solution arrived at in this way, as a limit of Bayes solutions, is called by Wald a Bayes solution in the wide sense. His theorem is, that under very general conditions the class of all Bayes solutions in the wide sense is a complete class.

The simple example we have taken does not indicate in the least the extreme degree of generality which attaches to this result of Wald's. Some idea of this generality can be obtained from the consideration that about half the present book is devoted to the statement of the conditions under which the theorem is true, and to its proof. In our example, the successive observations are statistically independent; this restriction is removed. Again, in our case, the number of observations was fixed in advance; this restriction is removed, and in its place very general assumptions are made to the effect that observations cost money. The possibility is allowed that observations of several different kinds can be made. In our case the problem reduced to the estimation of a single parameter; this restriction is removed, so that we need not even be estimating any finite number of parameters, and, if we are, we need not make point estimates of them. Finally, included in the decision function is not only the final decision to be arrived at, but also which observations are to be made at each stage of the experiment, and when experimentation is to be stopped, etc. It is, in fact, difficult to see in what direction any serious possibility of generalization remains. And it is, of course, precisely in its degree of generality that the value of the theorem lies. Its meaning can be summed up by saying that, if we are prepared to accept the principle of minimizing mean loss, when this is known, then we should use Bayes's theorem to guide us, based on some *a priori* distribution.

Thus we would conjecture that a major result of this work of Wald's will be to reinstate Bayes's theorem to the central place it always should have occupied in the theory of decision functions. It lost this place, probably, because of Fisher's attacks on its use in statistical inference. Fisher was quite right in asserting that Bayes's theorem had no application to the problems of statistical inference with which Fisher, and the majority of statisticians after him, have been concerned. But this was taken on many sides to mean that the theorem had no application to the problems, such as those of sampling inspection, which could rightly be regarded primarily as decision problems rather than inference problems.

Another theme which occurs in the present work is that of the 'minimax' decision function. One of the difficulties that has to be faced by anyone who tries to identify the decision problem and the inference problem is, that we have a feeling that the inference to be drawn from a given set of data in relation to a given set of statistical hypotheses should be unique (though it may be uncertain). And the trouble with Bayes solutions is that there are as many of them as there are prior distributions, in general (apart from cases, for example, where a uniformly most powerful test exists). So some principle has to be adopted which will pick out from the multiplicity of Bayes solutions one which can in some sense or other be regarded as the 'true' solution. Bayes and Laplace, of course, did this by their famous 'equal distribution of ignorance' postulate. Now Wald suggests, instead, the principle of 'minimizing the maximum loss'. This means choosing that decision function for which the maximum of the risk function is least. In the example we have taken, this means using the decision rule (iii). For we have already seen that this rule is a Bayes rule, and hence admissible. This means that no other rule can have a risk function which is never greater than the risk function for this rule. But since the risk function for rule (iii) is a constant, this means that any other risk function must have its maximum value greater than or equal to this constant. In other words, rule (iii) minimizes the maximum loss.

Minimax rules have useful properties, perhaps the most useful being, as in our example, that the risk function for such rules is usually constant. But the principle of adopting the minimax rule cannot be regarded as any less arbitrary than Bayes's and Laplace's principle of the equal distribution of the prior probability. If, in a long series of repetitions of the problem we have considered, *in fact* p has a distribution, concentrated near the ends of its range, then decision rule (i) will *in fact* be the best of those considered; and if *in fact* there is a concentration towards $\frac{1}{2}$, then perhaps (iii), or even (ii), will be better than (i). The prior distribution really is a feature of the problem, and without knowing it we *cannot* obtain a unique solution, which has any absolute title to being called the 'best'.

From the point of view of statisticians mainly concerned with problems of inductive inference, perhaps the most useful aspect of Wald's main result is, that in so far as every admissible rule is a Bayes rule, since the data enter into the posterior probability only via the likelihood function, this represents another way of seeing that the likelihood function 'contains all the information in the data', and hence that the problem of inference reduces to the problem of describing the likelihood function.

G. A. BARNARD

PUBLICATIONS OF U.S. DEPARTMENT OF COMMERCE, NATIONAL BUREAU OF STANDARDS

(i) **Table of arctangents of rational numbers.** Applied Mathematics Series 11. 1951. Pp. 105. Price $1.50.

This table gives arctangents for rational arguments m/n, i.e. it provides $\tan^{-1}(m/n)$ to 12 decimals for all pairs of integers m, n with $0 < m < n \leqslant 100$. Also shown are $\cot^{-1}(m/n) = \frac{1}{2}\pi - \tan^{-1}(m/n)$ (to 12 decimals) and the sum of squares m^2+n^2.

The main applications of the table are:

(*a*) The conversion of cartesian co-ordinates n, m to polars $\phi = \tan^{-1}(m/n)$ and $r = \sqrt{(m^2+n^2)}$ for which an auxiliary table of square roots would be required.

(*b*) The evaluation of natural logarithms for complex arguments from

$$\log_e(n+im) = \tfrac{1}{2}\log_e(m^2+n^2)+i\tan^{-1}(m/n)$$

for which an auxiliary table of natural logarithms would be required.

(*c*) The evaluation of the Γ-function for complex arguments from the recurrence

$$\log_e\Gamma(x+iy+1) = \tfrac{1}{2}\log_e(x^2+y^2)+i\tan^{-1}(y/x)+\log_e\Gamma(x+iy).$$

The latter property may make the table useful in statistical work in connection with moment generating functions of Γ-variables or Laguerre series.

Auxiliary tables provide formulae for the 'reduction' of arctangents in the forms

$$\tan^{-1}(m/n) = \Sigma f_j \tan^{-1} r_j \quad \text{(Table 1, col. 3)},$$
$$\tan^{-1} m \quad = \Sigma f_j \tan^{-1} m_j \quad \text{(Table 2)},$$

by providing the integers f_j, r_j and $m_j < m$. The table is provided with a short Introduction prepared by J. Todd and giving theoretical foundations and details of interpolation.

(ii) **Tables of the exponential function e^x.** Applied Mathematics Series 14 (supersedes MT 2). 1951. Pp. 537. Price $3.25.

This table constitutes a revised edition of the second volume of the series of mathematical tables published by the Work Projects Administration of the City of New York and was originally published in 1939. It represents the most extensive table of the exponential function available if one takes into consideration both the decimal accuracy in e^x and the fineness of the tabular interval in x.

It provides the following functions:

e^x to 18 decimals for	$x =$ 0·0000	(0·0001)	1·0000
to 15 decimals for	1·0000	(0·0001)	2·5000
to 15 decimals for	2·500	(0·001)	5·000
to 15 decimals for	5·00	(0·01)	10·00
e^{-x} to 18 decimals for	0·0000	(0·0001)	2·5000
e^x and e^{-x} to 18 decimals for	0	(10^{-6})	10^{-4}
to 19 significant figures for	1	(1)	100

and auxiliary tables.

It should be noted that e^x and e^{-x} are given in separate tables and, if both required for the same argument, two separate pages have to be consulted. A common argument for $x \leqslant 2{\cdot}5$ would have been preferable.

(iii) A guide to tables of the normal probability integral. Applied Mathematics Series 21. 1952. Pp. 16. Price 15 c.

This issue of the series provides, in Part I, a brief index of tables of the normal probability integral. In addition to the reference, information is provided on the type of area tabulated (i.e. whether single tail, two tail, etc.) the decimal accuracy, the tabular intervals and auxiliary quantities (e.g. differences) provided for interpolation. Lay-out and arrangement of this index are therefore similar to those adopted in *An Index of Mathematical Tables* by Fletcher, Miller & Rosenhead (Scientific Computing Service Ltd., London and McGraw-Hill Book Co., New York, 1946), although the present list of specialized tables is more extensive.

Part II gives the corresponding index for tables of the inverse function (normal deviates).

Part III gives principles of interpolation applicable to these tables in general and Part IV a summary of formulae.

(iv) Tables of the Bessel functions $Y_0(x)$, $Y_1(x)$, $K_0(x)$, $K_1(x)$. Applied Mathematics Series 25 (supersedes A.M.S.I.). 1952. Pp. 60. Price 40 c.

This table represents a subtabulation of the Bessel functions Y_0, Y_1; K_0, K_1 given at a rather wider interval in vol. VI of the mathematical tables series issued by the British Association for the Advancement of Science; certain recalculations were made for small x. The functions tabled are

$$Y_0(x) \text{ and } Y_1(x) \quad \text{for} \quad x = 0{\cdot}0001\,(0{\cdot}0001)\,0{\cdot}050\,(0{\cdot}001)\,1{\cdot}000$$
$$K_0(x) \text{ and } K_1(x) \quad \text{for} \quad x = 0{\cdot}0001\,(0{\cdot}0001)\,(0{\cdot}033)\,(0{\cdot}001)\,1{\cdot}000$$

to the same (varying) decimal accuracy as in the British Association tables with an additional interpolation tolerance of 1–2 units. For small x auxiliary functions are tabled which enable the Bessel functions to be evaluated with the help of log tables. An Introduction by A. N. Lowan gives basic formulae and methods of interpolation mainly based on the tabled differences.

H. O. HARTLEY.

Additional Note by Editor. Since these reviews were written an additional table in this series has been received, which is a re-issue of Table 16 of the N.B.S. Mathematical Tables Project.

(v) Table of arctan x. Applied Mathematics Series 26. 1952. Pp. 188. Price $1.75.

The integral

$$\operatorname{Arctan} x = \int_0^x \frac{du}{1+u^2}$$

is tabled to 12 decimal places for the following ranges and argument intervals:

$$x = 0{\cdot}000\,(0{\cdot}001)\,7{\cdot}000\,(0{\cdot}01)\,50{\cdot}00\,(0{\cdot}1)\,300{\cdot}0\,(1)\,2000\,(10)\,10{,}000.$$

Second differences are given throughout.